FORM PTO-1390 (REV. 12-2001)
U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE

TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371

ATTORNEY 'S DOCKET NUMBER
216019-67

U.S. APPLICATION NO. (If known, see 37 CFR 1.5
10/069025

INTERNATIONAL APPLICATION NO.	INTERNATIONAL FILING DATE	PRIORITY DATE CLAIMED
PCT/US00/22889	18 August 2000	21 August 1999

TITLE OF INVENTION
GENE EXPRESSION MODULATED IN ATAXIA TELANGIECTASIA TUMORGENESIS

APPLICANT(S) FOR DO/EO/US
BARLOW, et al.

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. [X] This is a **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. [ ] This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. [X] This is an express request to begin national examination procedures (35 U.S.C. 371(f)). The submission must include items (5), (6), (9) and (21) indicated below.
4. [ ] The US has been elected by the expiration of 19 months from the priority date (Article 31).
5. [X] A copy of the International Application as filed (35 U.S.C. 371(c)(2))
   - a. [ ] is attached hereto (required only if not communicated by the International Bureau).
   - b. [ ] has been communicated by the International Bureau.
   - c. [X] is not required, as the application was filed in the United States Receiving Office (RO/US).
6. [ ] An English language translation of the International Application as filed (35 U.S.C. 371(c)(2)).
   - a. [ ] is attached hereto.
   - b. [ ] has been previously submitted under 35 U.S.C. 154(d)(4).
7. [X] Amendments to the claims of the International Aplication under PCT Article 19 (35 U.S.C. 371(c)(3))
   - a. [ ] are attached hereto (required only if not communicated by the International Bureau).
   - b. [ ] have been communicated by the International Bureau.
   - c. [ ] have not been made; however, the time limit for making such amendments has NOT expired.
   - d. [X] have not been made and will not be made.
8. [ ] An English language translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371 (c)(3)).
9. [ ] An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).
10. [ ] An English lanugage translation of the annexes of the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

Items 11 to 20 below concern document(s) or information included:

11. [ ] An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
12. [ ] An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.
13. [X] A FIRST preliminary amendment.
14. [ ] A SECOND or SUBSEQUENT preliminary amendment.
15. [ ] A substitute specification.
16. [ ] A change of power of attorney and/or address letter.
17. [ ] A computer-readable form of the sequence listing in accordance with PCT Rule 13ter.2 and 35 U.S.C. 1.821 - 1.825.
18. [ ] A second copy of the published international application under 35 U.S.C. 154(d)(4).
19. [ ] A second copy of the English language translation of the international application under 35 U.S.C. 154(d)(4).
20. [ ] Other items or information:

JC10 Rec'd PCT/PTO 19 FEB 2002

U.S. APPLICATION NO. (if known, see 37 CFR 1.5)	INTERNATIONAL APPLICATION NO.	ATTORNEY'S DOCKET NUMBER
60/149,868 10/069025	PCT/US00/22889	216019-67

	CALCULATIONS	PTO USE ONLY
21. [X] The following fees are submitted: **BASIC NATIONAL FEE (37 CFR 1.492 (a) (1) - (5)):** Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO . . . . **$1040.00** International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO . . . . **$890.00** International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO . . . . **$740.00** International preliminary examination fee (37 CFR 1.482) paid to USPTO but all claims did not satisfy provisions of PCT Article 33(1)-(4) . . . . **$710.00** International preliminary examination fee (37 CFR 1.482) paid to USPTO and all claims satisfied provisions of PCT Article 33(1)-(4) **$100.00**		
ENTER APPROPRIATE BASIC FEE AMOUNT =	$ 740.00	
Surcharge of **$130.00** for furnishing the oath or declaration later than [ ] 20 [ ] 30 months from the earliest claimed priority date (37 CFR 1.492(e)).	$	

CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE	$	
Total claims	43 - 20 =	23	x **$18.00**	$ 414.00	
Independent claims	24 - 3 =	21	x **$84.00**	$1,764.00	
MULTIPLE DEPENDENT CLAIM(S) (if applicable)			+ **$280.00**	$	
TOTAL OF ABOVE CALCULATIONS =				$2,918.00	
[ ] Applicant claims small entity status. See 37 CFR 1.27. The fees indicated above are reduced by 1/2. +				$	
SUBTOTAL =				$2,918.00	
Processing fee of **$130.00** for furnishing the English translation later than [ ] 20 [ ] 30 months from the earliest claimed priority date (37 CFR 1.492(f)).				$	
TOTAL NATIONAL FEE =				$2,918.00	
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31). **$40.00** per property +				$	
TOTAL FEES ENCLOSED =				$2,918.00	
				Amount to be refunded:	$
				charged:	$

a. [X] A check in the amount of $ 2918.00 to cover the above fees is enclosed.

b. [ ] Please charge my Deposit Account No. ____________ in the amount of $ __________ to cover the above fees. A duplicate copy of this sheet is enclosed.

c. [X] The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 50-0639. A duplicate copy of this sheet is enclosed.

d. [ ] Fees are to be charged to a credit card. **WARNING:** Information on this form may become public. **Credit card information should not be included on this form.** Provide credit card information and authorization on PTO-2038.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137 (a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO:

Brian M. Berliner
O'Melveny & Myers LLP
400 South Hope Street
Los Angeles, CA 90071-2899

SIGNATURE

Brian M. Berliner
NAME

34,549
REGISTRATION NUMBER

10/069025

JC10 Rec'd PCT/PTO 19 FEB 2002

PATENT
216,019-67

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants: BARLOW et al.	
Serial No.:	Art Unit:
National Stage Filing Date:	
International Appln. No: PCT/US00/22889	Examiner:
International Filing Date: August 18, 2000	
Title: Gene Expression Modulated in Ataxia Telangiectasia Tumorgenesis	

PRELIMINARY AMENDMENT

Box PCT
Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

Preliminary to examination, please amend the above-identified patent application as follows:

IN THE SPECIFICATION:

Please amend the specification as follows:

Please replace the paragraph beginning on page 61, line 14, with the following rewritten paragraph:

In general, double-stranded cDNA is generated from poly(A)-enriched cytoplasmic RNA extracted from the tissue samples of interest using an equimolar mixture of all 48 5'-biotinylated anchor primers of a set to initiate reverse transcription. One such suitable set is G-A-A-T-T-C-A-A-C-T-G-G-A-A-G-C-G-G-C-C-G-C-A-G-G-A-A-T-T-T-T-T-T-T-T-T-T-T-T-T-T-T-T-T-V-N-N (SEQ ID NO:24), where V is A, C or G and N is A, C, G or T. One member of this mixture of 48 anchor primers initiates synthesis at a fixed position at the 3' end of all copies of each mRNA species in the sample, thereby defining a 3' endpoint for each species, resulting in biotinylated double stranded cDNA.

IN THE SEQUENCE LISTING:

Also attached is an amended version of the entire Sequence Listing for this application, submitted in both written and electronic format. The amendments made to the sequence listing effect the following changes: (1) on page 5 of the sequence listing, on the 31st line containing typed text, replace "<211> 49" with "<211> 48"; (2) on page 5 of the sequence listing, on the 36th line containing typed text, change "<222> (48) . . (49)" to "<222> (47) . . (48)"; (3) on page 5 of the sequence listing on the 40th line containing typed text, change "<222> (47)" to "<222> (46)"; (4) on page 5 of the sequence listing on the 45th line containing typed text, change "gaattcaact ggaagcggcc cgcaggaatt tttttttttt ttttttvnn 49" to "gaattcaact ggaagcggcc gcaggaattt tttttttttt tttttvnn 48". To effect the foregoing changes, please replace the sequence listing in its entirety with the amended sequence listing enclosed herein. We have submitted the entire

sequence listing in written and electronic format for replacement because the written and electronic sequence listing submitted herein are in the PatentIn 3.1 format, and may differ from the PatentIn 2.1 version originally submitted to the U.S. Receiving Office in the international stage of this application due to differences in the two software versions, which are primarily formatting in nature. The differences we noted were as follows:

1) The line spacing for the sequences is double spaced for PatentIn 3.1, whereas for PatentIn 2.1 it was single spaced. Thus the overall document has more pages with the newer version of PatentIn.

2) In the filed application, blanks appeared for fields <140> and <141> (they are in the header information). With the PatentIn 3.1 version, these field numbers do not appear at all.

3) Notes for the primer sequences (SEQ ID NO: 24, 26, and 28) appear in a different order when PatentIn 3.1 is used. For the sequence listing generated with PatentIn 3.1, the Description line preceeds the misc_feature line. In the PatentIn 2.1, this was reversed.

4) Miscellaneous features that are 1 base pair long (see SEQ ID NO: 24) are listed with the same base pair number as the beginning and end nucleotide number (if PatentIn 3.1 is used) rather than with one number (possible with PatentIn 2.1).

REMARKS

The specification and sequence listing are amended as described above. It is noted that the amendments to the specification refer to page and line numbers as set forth in the originally filed English version of the application. No new matter has been added to the application by this preliminary amendment, and this submission does not include matter that goes beyond the disclosure in the international application as filed. The sequence listing recorded and submitted herewith in computer readable form is identical to the written sequence listing submitted herewith, and is substantially identical

to the sequence listing submitted previously (except for the amendment described above and version differences noted above). Accordingly, the Applicant respectfully requests an examination of the application on the merits in view of the foregoing amendments.

These amendments result from a typographical error that was recently noticed, i.e., an extra "C" had been introduced into the Not I site of one set of anchor primers, SEQ ID NO:24. The amendments correct an obvious error, and no new matter is added. The text of the application makes it clear that the anchor primer contains a Not I site. One of ordinary skill would understand that the sequence of the Not I site is a palindrome, G-C-G-G-C-C-G-C. The typographical error, G-C-G-G-C-C-C-G-C, would be recognized as an obvious error because it is not a palindrome, and not the Not I recognition site sequence. See page 48 of the 1998/1999 New England Biolabs Catalog, attached as Exhibit 1.

Attached hereto is a marked-up version of the changes made to the specification by the current amendment. The attached page is captioned "**Version with markings to show changes made**."

Respectfully submitted,

Date: February 19, 2002

Brian M. Berliner
Attorney for Applicants
Registration No. 34,549

O'MELVENY & MYERS LLP
400 So. Hope Street
Los Angeles, California 90071-2899
Telephone: 213-430-6000

Enclosures: (1) New England Biolabs Catalog Pages
(2) Amended Sequence Listing (written)
(3) Amended Sequence Listing (diskette, PatentIn 3.1)

LA1:961207.1

VERSION WITH MARKINGS TO SHOW CHANGES MADE

IN THE SPECIFICATION:

The paragraph beginning on page 61, line 14 is amended as follows:

In general, double-stranded cDNA is generated from poly(A)-enriched cytoplasmic RNA extracted from the tissue samples of interest using an equimolar mixture of all 48 5'-biotinylated anchor primers of a set to initiate reverse transcription. One such suitable set is G-A-A-T-T-C-A-A-C-T-G-G-A-A-G-C-G-G-C-C-**[C-]**G-C-A-G-G-A-A-T-T-T-T-T-T-T-T-T-T-T-T-T-T-T-T-T-T-V-N-N (SEQ ID NO:63), where V is A, C or G and N is A, C, G or T. One member of this mixture of 48 anchor primers initiates synthesis at a fixed position at the 3' end of all copies of each mRNA species in the sample, thereby defining a 3' endpoint for each species, resulting in biotinylated double stranded cDNA.

SEQUENCE LISTING

<110> Barlow, Carrolee
Winrow, Christopher J
Callahan, Marie Lei A
Pankratz, Daniel G

<120> Gene Expression Modulated in Ataxia Telangiectasia Tumorgenesis

<130> 99-529B

<140>
<141>

<150> 60/149,868
<151> 1999-08-19

<160> 52

<170> PatentIn Ver. 2.1

<210> 1
<211> 406
<212> DNA
<213> Mus musculus

<400> 1
```
cggacgggga attcccaaaa acactgcacg aagttaagct gacagtacag aaggatcagg 60
tgtgtgagtc ccagttccaa agttcttaca acagagctaa tgagatatgt gtgggagact 120
caaagatcaa gggagcttcc tttgaggagg attctggagg cccgcttgtg tgtaaaagag 180
cagctgcagg catcgtctcc tacgggcaaa ctgatggatc agctccacaa gtcttcacaa 240
gagttttggg ttttgtatcg tggataaaga aaacgatgaa acacagctaa ctacaagaag 300
caaactagat cctgatctga ccagccatct tccccatagc tgagtccagg attgctctag 360
gacagatggc agcaactgaa taaagaactt tctctgactc caaaaa                406
```

<210> 2
<211> 54
<212> DNA
<213> Mus musculus

<400> 2
```
cggcagaggc tcctggcctg cacttttcct attaaagtca ctgacacaca aaaa       54
```

<210> 3
<211> 101
<212> DNA
<213> Mus musculus

<400> 3
```
cggccgtgtg tgccttagga gtgggaaact ttgcatttct ctctccttat ccttcttgta 60
agacatccat ttaataaagt ctcatgctga gagagtaaaa a                     101
```

<210> 4
<211> 258
<212> DNA
<213> Mus musculus

<400> 4
cggctcgcag agtgaaggca cccagccatc aggaatgcag ccctgccagc caggcaccat 60
gggtggcagc catcatgctt tcatgaccat tgggcctgct ggtctacctg gtcttagccc 120
aggaagccac caggtaggtt agggtggtca gtgccgagtc tcctgccaga cacagttata 180
cctgcctttc tgcactgctc caggcatgcc cttagagcat ggtgttttaa agctaaataa 240
agtctctaac tacaaaaa 258

<210> 5
<211> 116
<212> DNA
<213> Mus musculus

<400> 5
cgggctcagc actgtctgct cacgtcccca ctgttgcaac acctagtctg tttgcaagga 60
ggacaggaca ggtcacatgg agctttatgt caataaagtc tttatcttgt caaaaa 116

<210> 6
<211> 68
<212> DNA
<213> Mus musculus

<400> 6
cggaacgggg gatgatgggg gattggaaga acttgaactt tttattaaaa tattgtaagt 60
cctaaaaa 68

<210> 7
<211> 172
<212> DNA
<213> Mus musculus

<400> 7
cggaccgcca tgcctggtga gcatatccat cctttgaaag atacactcat gtcctaaacg 60
gagctgtgac tgagatttag tgttttcaat atgaacaaaa agcatttgcc tagcaaatta 120
agttcaaatg aaggcttagt cttttgaaga ataaagcata aaaccatcaa aa 172

<210> 8
<211> 151
<212> DNA
<213> Mus musculus

<400> 8
cggacgctgg agccttcctt aatccacacc tttgaagcca attctcaact ggttgccgtg 60
tacactgatt acatttcagg aggacgtgtt ctatttatgt tgtgcctctg caggcaaagc 120
ccttaataaa tattttgcat actttcaaaa a 151

<210> 9
<211> 300
<212> DNA
<213> Mus musculus

<400> 9
cggacggtgt accccgagga tcgccccagg tggagggaaa gatccaggac caggtcgcgc 60
agcaggagta gaaccccatt tcgcctgtgt gagaaagatc gaatggagct actagaaata 120
gcaaaagcca acgcagcaaa agctctggga acagccaact tcgacttgcc agcaagtctc 180
cgagccaagg aggcaagcca ggggacagct gtttccagca gtgggccaaa ggtggagcat 240
tcagaaaagc agactgaaga tacaactaaa aataccagtg aaaagtcttc tacacaaaaa 300

<210> 10
<211> 43
<212> DNA
<213> Mus musculus

<400> 10
cggagtcccc tcctctgaat taaaatcaca ctgagcggaa aaa 43

<210> 11
<211> 400
<212> DNA
<213> Mus musculus

<400> 11
cggcatcgtc cttgttaagc actgctttcc agaccgtgtc tgacttagtc aacctagagg 60
attttatagt ttcactctaa tgaagtcctg attaacacag agcagcggag cagcggtccc 120
tcccctgagt ggcagcagca ggaagccccg tttgtaatgt gcgttgcgct gctttaccgt 180
acacccagtg gcattctttt tactataacg gttttccccc ctctcagaag agaaatctta 240
aaatttgcaa cagacctcct aacttttgtt tacttttttt gtcttaacga tagatttgaa 300
aatgaaaaat gttaaaaaga caactggatc tgttgtccgt attatatttg caatctgaaa 360
tttgtgtgaa ttgatttagt aaaatgttaa actggaaaaa 400

<210> 12
<211> 219
<212> DNA
<213> Mus musculus

<400> 12
cggcctggac cacatcacgg aggaagaact gaaacggttc aagtactttg ctttgactga 60
gtttcagatt gccaggagca cactcgacgt cgcagatagg acagaattag ctgaccacct 120
gattcaaagt gcatgtgcgg cgtctggagt gaccaaggcc attaatattt tccagaagtt 180
gaattatatg catattgcaa atgctcttga agagaaaaa 219

<210> 13
<211> 89
<212> DNA
<213> Mus musculus

<400> 13
cgggctcttt gcacagttct aacctataca acttggtctg ctcctcctac caacttattt 60
aatagacatt aaagtaaaac gcccaaaaa 89

<210> 14
<211> 42
<212> DNA
<213> Mus musculus

<400> 14
cggctatcag tgctaacaga aaacattgta gaaggccaaa aa 42

<210> 15
<211> 247
<212> DNA
<213> Mus musculus

<400> 15

cggaccatcc ctgctctgtg tccatctact catccaatgg agtcattagc agtcaatcgc 60
agcttcactg gacactgagg ggtcacagac ttaggctccc tttgggtcac gtccagcgtg 120
tcctagactt tatcatcttt cagaggcggt aggcagactg ttcacaaagg ctttctctag 180
ctttccactg caattaatct tggtcactcc ctcaaataag tatatttttt ctagaaaagg 240
ggaaaaa 247

<210> 16
<211> 260
<212> DNA
<213> Mus musculus

<400> 16
cggaccctga ttcgcatcat ggtgtctcgc ggcgagcttg acctgctgga tatccgagca 60
gagtataagc ggatgtacgg caagtcactg taccacgata tcacgggaga tacttctgga 120
gactatcgaa agattccact gaagatctgt ggtggcaacg actgagctgt ggcagtggct 180
cacctgtgcc ccctgctggc attggccagg aaacgtccca aagagcacac ttcggtttcc 240
ctatctaata accccaaaaa 260

<210> 17
<211> 209
<212> DNA
<213> Mus musculus

<400> 17
cggaccgtgc aaaggtagca taatcacttg ttccttaatt agggactagc atgaacggct 60
aaacgagggt ccaactgtct cttatcttta atcagtgaaa ttgacctttc agtgaagagg 120
ctgaaatata ataataagac gagaagaccc tatggggctt aaattatata acttatctat 180
ttaatttatt aaacctaatg gcccaaaaa 209

<210> 18
<211> 385
<212> DNA
<213> Mus musculus

<400> 18
cggagtggac agcacaagcc atggaagcgc agatggccca gtgagccctc agcgctgtgc 60
gcactgcatc tttgccactt cgtgcagatt gattcactgc ccctcaaagc atttgcatca 120
tgaccttaga catctcagcc tcctttgctg gatcccactt gaagaagatg ctagagcaag 180
aagtgcatcg tttaaacata atagctcccg tgagatctgg gtctgtgtat tgaactaact 240
cagacgcttt gaaaacttgt tctttaagtg gaagtgaaag acctgaagtt tgggatttgt 300
gatcactttg tagtcatgta aaacttaagt gctttgtgcc tctccacatg tggttactct 360
aataaacgga gcggtgagcc aaaaa 385

<210> 19
<211> 83
<212> DNA
<213> Mus musculus

<400> 19
cggccgaggc gaggcgccgc gcggaaaacc gcggcgggtg ttgacgcgat gtgatttctg 60
cccagtgctc tgaatgtcaa aaa 83

<210> 20
<211> 286
<212> DNA
<213> Mus musculus

<400> 20
cgggggccaa gacgaggagg tgctgatgat gaacggtcat cctggcggaa tgctgatgat 60
gataggggtc ccaggagagg catggatgat gataggggtc ccaggcgagg gctggatgat 120
gaccgaggac cttggaggaa tgctgctgaa gataggattt ccaggagagg tgcagatgat 180
gaccgagggc cttggagaaa tatggatgat gatagggttc ctagaagggg tgatgatgca 240
agacctggtc cttggagatc atttgtcgag ccaggtggat gaaaaa 286

<210> 21
<211> 124
<212> DNA
<213> Mus musculus

<400> 21
cggtaccaat gtatagctgt gttggagaaa tatggaagat aaactataca ggatatttgc 60
acattctcat ctgtctctct ggtgaccgat aattcagttt gctccagctg cccaagccca 120
aaaa 124

<210> 22
<211> 68
<212> DNA
<213> Mus musculus

<400> 22
cggtagcggg aacaccgaga acgaagccag cctggctttc gaaaacgcct gggtttgctg 60
gagaaaaa 68

<210> 23
<211> 160
<212> DNA
<213> Mus musculus

<400> 23
cggtgcagca ccgtgaggtc cagtaccacg agtccgacac cttccgtggc tacttcaagc 60
agggcatcat ctataaaaag gggggtgtgg cctctgggat gaagcacgtg gagaccaaca 120
cctatgatgt gaagaggctg ctacacgtaa aggggaaaaa 160

<210> 24
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> (48)..(49)
<223> N stands for any nucleotide.

<220>
<221> misc_feature
<222> (47)
<223> V stands for A, G, or C.

<220>
<223> Description of Artificial Sequence: synthetic primer

<400> 24
gaattcaact ggaagcggcc cgcaggaatt tttttttttt ttttttvnn 49

<210> 25
<211> 16
<212> DNA
<213> Artificial Sequence

<223> Description of Artificial Sequence: synthetic
primer

<400> 25
aggtcgacgg tatcgg 16

<210> 26
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> (16)
<223> N stands for any nucleotide.

<223> Description of Artificial Sequence: synthetic
primer

<400> 26
ggtcgacggt atcggn 16

<210> 27
<211> 15
<212> DNA
<213> Artificial Sequence

<223> Description of Artificial Sequence: synthetic
primer

<400> 27
gagctccacc gcggt 15

<210> 28
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> (13)..(16)
<223> N stands for any nucleotide.

<223> Description of Artificial Sequence: synthetic
primer

<400> 28
cgacggtatc ggnnnn 16

<210> 29
<211> 16
<212> DNA
<213> Artificial Sequence

<223> Description of Artificial Sequence: synthetic
primer

<400> 29
cgacggtatc gggagc 16

<210> 30
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1_7.

<400> 30
gatcgaatcc ggacggtgta ccccgaggat 30

<210> 31
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1_8.

<400> 31
gatcgaatcc gggctctttg cacagttcta 30

<210> 32
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1_11.

<400> 32
gatcgaatcc ggccgtgtgt gccttaggag 30

<210> 33
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1_16.

<400> 33
gatcgaatcc ggaacggggg atgatggggg 30

<210> 34
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1_20.

<400> 34
gatcgaatcc ggtaccaatg tatagctgtg 30

<210> 35
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1_21.

<400> 35
gatcgaatcc ggcagaggct cctggcctgc 30

<210> 36
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1_22.

<400> 36
gatcgaatcc ggctatcagt gctaacagaa 30

<210> 37
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1_24.

<400> 37
gatcgaatcc gggctcagca ctgtctgctc 30

<210> 38
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1_27.

<400> 38
gatcgaatcc ggacggggaa ttcccaaaaa 30

<210> 39
<211> 30
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA
primer for BAR1_28.

<400> 39
gatcgaatcc ggctcgcaga gtgaaggcac 30

<210> 40
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA
primer for BAR1_29.

<400> 40
gatcgaatcc ggagtcccct cctctgaatt 30

<210> 41
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA
primer for BAR1_34.

<400> 41
gatcgaatcc ggaccatccc tgctctgtgt 30

<210> 42
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA
primer for BAR1_35.

<400> 42
gatcgaatcc ggaccctgat tcgcatcatg 30

<210> 43
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA
primer for BAR1_37.

<400> 43
gatcgaatcc ggaccgccat gcctggtgag 30

<210> 44
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA
primer for BAR1_38.

<400> 44
gatcgaatcc ggaccgtgca aaggtagcat 30

<210> 45
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA
primer for BAR1_39.

<400> 45
gatcgaatcc ggacgctgga gccttcctta 30

<210> 46
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA
primer for BAR1_41.

<400> 46
gatcgaatcc ggagtggaca gcacaagcca 30

<210> 47
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA
primer for BAR1_49.

<400> 47
gatcgaatcc ggcatcgtcc ttgttaagca 30

<210> 48
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA
primer for BAR1_50.

<400> 48

gatcgaatcc ggccgaggcg aggcgccgcg 30

<210> 49
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1_52.

<400> 49
gatcgaatcc ggcctggacc acatcacgga 30

<210> 50
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1_62.

<400> 50
gatcgaatcc gggggccaag acgaggaggt 30

<210> 51
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1_67.

<400> 51
gatcgaatcc ggtagcggga acaccgagaa 30

<210> 52
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1_69.

<400> 52
gatcgaatcc ggtgcagcac cgtgaggtcc 30

10/069025

Rec'd PCT/PTO 19 FEB 2002

WO 01/12647

PCT/US00/22889

GENE EXPRESSION MODULATED IN ATAXIA TELANGIECTASIA TUMORGENESIS

BACKGROUND OF THE INVENTION

Ataxia telangiectasia (AT) is a human autosomal recessive multisystem disorder comprising progressive cerebellar ataxia with onset in infancy, progressive oculocutaneous telangiectasia, and unusual susceptibility to progressive bronchopulmonary disease and to lymphoreticular neoplasia. See, generally, Boder, E., (1987) Ataxia-Telangiectasia, pp. 95-117 in Gomez, M.R. and Adams, R.D., Eds., Neurocutaneous Disease. A Practical Approach, Butterworths, Boston. In addition, AT generally includes other progressive neurologic degenerations such as choreoathetosis and oculomotor dysfunction, recurrent sinopulmonary infections secondary to immunodeficiency, lymphoreticular malignancies, growth retardation, incomplete sexual maturation, endocrine abnormalities, and premature aging of the skin and hair. Barlow, C. et al., (1996) *Atm*-deficient mice: a paradigm of ataxia telangiectasia, *Cell*, 86:159-171. The disease is progressive, and death generally occurs by the second or third decade of life due to neurologic deterioration or lymphoreticular malignancies, which occur in about 10% - 15% of patients. Taylor, A.M. et al. (1996) Leukemia and lymphoma in ataxia telangiectasia, *Blood*, 87:423-438. No effective treatments have been found to alter the course of the disease. The frequency of AT in the United States and Britain has been estimated to be between 1:40,000 and 1:100,000, resulting in a carrier frequency of 0.5% - 1.0% (Barlow et al., 1996). Heterozygous carriers may also have a predisposition to cancer, particularly breast cancer.

AT has been mapped to human chromosome position 11q22-23, and the responsible gene (designated *ATM*, "AT mutated") has been has been defined by positional cloning. The mouse homolog, *Atm*, has also been identified. *Atm* shows an 84% amino acid identity and a 91% similarity with *ATM*. Sequence comparisons revealed that both *ATM* and *Atm* are members of a family of genes involved at several stages in double-stranded DNA breakage repair and cell cycle control.

The transformation of a normal cell to a tumor cell appears to depend in part on mutations in genes that normally control the cell cycle. Cell cycle checkpoints are believed to play a major role in maintaining the integrity of the genome. Defects in these control points

may contribute to increased incidence of genomic alterations such as deletions, translocations, and amplifications that are common during the evolution of a normal cell to a cancer cell.

ATM is a key protein for managing cell cycle perturbations in response to DNA damage and plays a role in genetic stability and cancer susceptibility. Inactivation of normal ATM function could thus result in inappropriate replication of damaged DNA, thus contributing to the malignant transformation of cells. Available evidence indicates that ATM may be a critical regulator of many important cellular processes and thus has potential implications for cancer in the general population. Notably, many studies have indicated that loss of ATM function is related to cancer predisposition in AT heterozygotes. In particular, it has been found that women heterozygous for one mutated allele of *ATM* have a three to fivefold increase in breast cancer risk. Importantly, it has also been recently established that patients with loss of heterozygosity at 11q22 (the chromosomal location of *ATM*) and specifically have loss of ATM function, develop particularly aggressive forms of both B and T cell lymphomas. These studies suggest that *ATM* is a tumor suppressor gene whose inactivation is a key event in the development of many forms of cancer, particularly in T-cell prolymphocytic leukemias and aggressive forms of B cell chronic lymphocytic leukemia in patients who do not have AT.

SUMMARY OF THE INVENTION

The PCR-based Total Gene Expression Analysis (TOGA™) differential display system has been used in studies to examine how in ataxia telangiectasia tumorgenesis is regulated. Molecules have been identified that correspond to genes that are regulated in ataxia telangiectasia tumorgenesis. Such molecules are useful in the therapy of cancer, including treatment, prevention, and amelioration of cancer, and in the diagnosis of cancer.

The present invention provides novel polynucleotides and the encoded polypeptides. Moreover, the present invention relates to vectors, host cells, antibodies, and recombinant methods for producing the polynucleotides and the polypeptides. Also provided are diagnostic methods for detecting disorders related to the polypeptides and the polynucleotides encoding them, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying binding partners of the polypeptides.

In particular, it has been found that because of the close correspondence between human *ATM* and mouse *Atm* genes, the tumors that develop in *Atm*-deficient mice provide the ability to

specifically characterize the genes involved in cancer formation as a model of particular forms of human hematopoietic cancer. Cancer cell lines established from several T-cell lymphomas from multiple different *Atm* deficient animals are useful for the study of human cancers, as well as for the development of suitable diagnostic and therapeutic compositions and methods. An extensive cellular analysis of these tumors has been performed and show that they invariably arise at a stage of development when the TCRα undergoes rearrangement. It has been specifically found that the translocations occur at the TCRα locus in ten of ten tumor cell lines established from these mice. Therefore, it is likely that the tumors arise due to aberrant V(D)J rearrangements of the TCRα locus as seen in human cancers. In addition, it has been specifically found that regions of chromosome 12 are also involved. These regions are syntenic to the regions on human chromosomes known to be frequently mutated in human hematopoietic cancer. The *Atm* deficient tumor cell lines disclosed herein are useful for characterizing the genes involved in the multi-step process of cancer formation.

One embodiment of the invention provides an isolated nucleic acid molecule comprising a polynucleotide chosen from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, and SEQ ID NO:23. Further provided is a an isolated nucleic acid molecule comprising a polynucleotide at least 95% identical to the isolated nucleic acid molecule comprising a polynucleotide chosen from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, and SEQ ID NO:23. Also provided is an isolated nucleic acid molecule at least ten bases in length that is hybridizable to an isolated nucleic acid molecule selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, and SEQ ID NO:23 under stringent conditions. Preferably, an isolated nucleic acid molecule of the invention encodes a polypeptide epitope. A polynucleotide of the invention can further comprise sequential nucleotide deletions from either

the 5'-terminus or the 3'-terminus. A polynucleotide of the invention can be present in a recombinant vector and in a recombinant host cell.

Another embodiment of the invention provides an isolated polypeptide encoded by a polynucleotide chosen from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, and SEQ ID NO:23. Further provided is an isolated nucleic acid molecule encoding these polypeptides. Also provided is an isolated nucleic acid molecule encoding a fragment of a polypeptide encoded by a polynucleotide chosen from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, and SEQ ID NO:23. Additionally provided is an isolated nucleic acid encoding a species homologue of a polypeptide of the invention. Preferably, a polypeptide of the invention has biological activity. A polypeptide of the invention can comprise sequential amino acid deletions from either the C-terminus or the N-terminus.

Still another embodiment of the invention provides an isolated antibody that binds specifically to an isolated polypeptide of the invention. The antibody can be a monoclonal antibody or a polyclonal antibody.

Yet another embodiment of the invention provides an isolated polypeptide. The polypeptide is produced by culturing a recombinant host cell that comprises a polynucleotide of the invention under conditions such that the polypeptide is expressed and isolating the polypeptide.

Another embodiment of the invention provides a method for preventing, treating, modulating, or ameliorating a medical condition, such as ataxia telangiectasia, comprising administering to a mammalian subject a therapeutically effective amount of a polypeptide of the invention or a polynucleotide of the invention.

A further embodiment of the invention provides an isolated antibody that binds specifically to the isolated polypeptide of the invention. A preferred embodiment of the invention provides a method for preventing, treating, modulating, or ameliorating a medical condition, such as ataxia telangiectasia, comprising administering to a mammalian subject a therapeutically effective amount of the antibody.

An additional embodiment of the invention provides a method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject. The method comprises determining the presence or absence of a mutation in a polynucleotide of the invention. A pathological condition or a susceptibility to a pathological condition, such as ataxia telangiectasia is diagnosed based on the presence or absence of the mutation.

Even another embodiment of the invention provides a method of diagnosing a pathological condition or a susceptibility to a pathological condition, such as ataxia telangiectasia in a subject. The method comprises detecting an alteration in expression of a polypeptide encoded by the polynucleotide of the invention, wherein the presence of an alteration in expression of the polypeptide is indicative of the pathological condition or susceptibility to the pathological condition. The alteration in expression can be an increase in the amount of expression or a decrease in the amount of expression. In a preferred embodiment a first biological sample is obtained from a patient suspected of having ataxia telangiectasia and a second sample from a suitable comparable control source is obtained. The amount of at least one polypeptide encoded by a polynucleotide of the invention is determined in the first and second sample. The amount of the polypeptide in the first and second samples is determined. A patient is diagnosed as having ataxia telangiectasia if the amount of the polypeptide in the first sample is greater than or less than the amount of the polypeptide in the second sample.

Yet another embodiment of the invention provides a method for identifying a binding partner to a polypeptide of the invention. The polypeptide is contacted with a binding partner. Whether the binding partner effects an activity of the polypeptide is determined.

Even another embodiment of the invention provides a method of identifying an activity of an expressed polypeptide in a biological assay. A polypeptide of the invention is expressed in a cell. The expressed polypeptide is isolated and tested for an activity in a biological assay. The activity of the expressed polypeptide is identified based on the test results.

Another embodiment of the invention provides a substantially pure isolated DNA molecule suitable for use as a probe for genes regulated in ataxia telangiectasia. The DNA molecule is chosen from the group consisting of the DNA molecules identified in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, and SEQ ID NO:23.

Still another embodiment of the invention provides a kit for detecting the presence of a polypeptide of the invention in a mammalian tissue sample comprising a first antibody which immunoreacts with a mammalian protein encoded by a gene corresponding to a polynucleotide or polypeptide of the invention in an amount sufficient for at least one assay and suitable packaging material. The kit can further comprise a labeled or unlabeled second antibody that binds to the first antibody. A label can comprise enzymes, radioisotopes, fluorescent compounds, colloidal metals, chemiluminescent compounds, phosphorescent compounds, or bioluminescent compounds.

Yet another embodiment of the invention provides a kit for detecting the presence of a genes encoding an protein comprising a polynucleotide of the invention, or fragment thereof having at least 10 contiguous bases, in an amount sufficient for at least one assay, and suitable packaging material.

Even another embodiment of the invention provides a method for detecting the presence of a nucleic acid encoding a protein in a mammalian tissue sample. A polynucleotide of the invention or fragment thereof having at least 10 contiguous bases is hybridized with the nucleic acid of the sample. The presence of the hybridization product is detected.

BRIEF DESCRIPTION OF THE DRAWINGS

These and other features, aspects, and advantages of the present invention will become better understood with reference to the following description, appended claims, and accompanying drawing where:

Figure 1 is a graphical representation of the results of TOGA analysis using a 5' PCR primer with parsing bases GAGC, showing PCR products produced from mRNA extracted from cell lines AT-4 (Fig. 1A), AT-7 (Fig. 1B), AT-12 (Fig. 1C), and AT-13 (Fig. 1D); the arrow indicates the vertical cursor connecting the position of peaks of length 197 base pairs;

Figure 2 is a graphical representation of the results of Northern Blot analysis using radiolabeled probe derived from a candidate gene (U88908) corresponding to GAGC 197, showing, hybridization with a band of mRNA at 4.2 kb that is expressed at various levels in cell lines AT-4, AT-7, AT-10, AT-11, AT-13; APT-3, P53-1, 101-7, 292-3, and thymus, but not detectable in cell line AT-12; the image below shows the methylene blue stained gel for quantification of mRNA loading;

Figure 3 is a graphical representation of the results of TOGA analysis using a 5' PCR primer with parsing bases GCTG, showing PCR products produced from mRNA extracted from cell lines AT-4 (Fig. 3A), AT-7 (Fig. 3B), AT-12 (Fig. 3C), and AT-13 (Fig. 3D); the arrow indicates the vertical cursor connecting the position of peaks of length 345 base pairs;

Figure 4 is a graphical representation of the results of Northern Blot analysis using radiolabeled probe derived from a candidate gene (Z31664) corresponding to GCTG 345, showing, hybridization with bands of mRNA at 4.0 kb and 3.7 kb that are expressed at various levels in cell lines AT-4, AT-11, AT-12, AT-13, 292-3, and polyA (pA) thymus mRNA, but not detectable in cell lines AT-7, AT-10, APT-3, P53-1, 101-7; the image below shows the methylene blue stained gel for quantification of mRNA loading;

Figure 5 is a graphical representation of the results of TOGA analysis using a 5' PCR primer with parsing bases ACAT, showing PCR products produced from mRNA extracted from cell lines AT-4 (Fig. 5A), AT-7 (Fig. 5B), AT-12 (Fig. 5C), and AT-13 (Fig. 5D); the arrow indicates the vertical cursor connecting the position of peaks of length 367 base pairs;

Figure 6 is a graphical representation of the results of Northern Blot analysis using radiolabeled probe derived from a candidate gene (U38252) corresponding to ACAT 367, showing, hybridization with a band of mRNA at 4.1 kb that is expressed at various levels in cell lines AT-7, AT-10, AT-12, AT-13 and APT-3, but not detectable in cell lines AT-4, AT-11,

P53-1, 101-7, 292-3 and polyA (pA) thymus mRNA; the image below shows the methylene blue stained gel for quantification of mRNA loading;

Figure 7 is a graphical representation of the results of TOGA analysis using a 5' PCR primer with parsing bases ACGG, showing PCR products produced from mRNA extracted from cell lines AT-4 (Fig. 7A), AT-7 (Fig. 7B), AT-12 (Fig. 7C), and AT-13 (Fig. 7D); the arrow indicates the vertical cursor connecting the position of peaks of length 458 base pairs;

Figure 8 is a graphical representation of the results of Northern Blot analysis using radiolabeled probe derived from a candidate gene (X12822) corresponding to ACGG 458, showing, hybridization with bands of mRNA at 1.7 kb and 1.1 kb that are expressed at various levels in cell lines AT-7 and AT-10, but not detectable in cell lines AT-4, AT-12, AT-11, AT-13; APT-3, P53-1, 101-7, and 292-3, and polyA (pA) thymus (only 1.1 kb) mRNA; the image below shows the methylene blue stained gel for quantification of mRNA loading;

Figure 9 is a graphical representation of the results of TOGA analysis using a 5' PCR primer with parsing bases CCGT, showing PCR products produced from mRNA extracted from cell lines AT-4 (Fig. 9A), AT-7 (Fig. 9B), AT-12 (Fig. 9C), and AT-13 (Fig. 9D); the arrow indicates the vertical cursor connecting the position of peaks of length 151 base pairs; and

Figure 10 is a graphical representation of a more detailed analysis of the 151 b.p. PCR product indicated in Figure 9, using the extended TOGA primer G-A-T-C-G-A-A-T-C-C-G-G-C-C-G-T-G-T-G-T-G-C-C-T-T-A-G-G-A-G (SEQ ID NO:32).

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Definitions

The following definitions are provided to facilitate understanding of certain terms used throughout this specification.

In the present invention, "isolated" refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring), and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be

"isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide.

In the present invention, a "secreted" protein refers to those proteins capable of being directed to the ER, secretory vesicles, or the extracellular space as a result of a signal sequence, as well as those proteins released into the extracellular space without necessarily containing a signal sequence. If the secreted protein is released into the extracellular space, the secreted protein can undergo extracellular processing to produce a "mature" protein. Release into the extracellular space can occur by many mechanisms, including exocytosis and proteolytic cleavage.

As used herein, a "polynucleotide" refers to a molecule having a nucleic acid sequence shown in SEQ ID NOs:1-23. For example, the polynucleotide can contain all or part of the nucleotide sequence of the full length cDNA sequence, including the 5' and 3' untranslated sequences, the coding region, with or without the signal sequence, the secreted protein coding region, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. Moreover, as used herein, a "polypeptide" refers to a molecule having the translated amino acid sequence generated from the polynucleotide as broadly defined.

A "polynucleotide" of the present invention also includes those polynucleotides capable of hybridizing, under stringent hybridization conditions, to sequences contained in SEQ ID NOs:1-23, or the complement thereof, or the cDNA. "Stringent hybridization conditions" refers to an overnight incubation at 42°C in a solution comprising 50% formamide, 5X SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5X Denhardt's solution, 10% dextran sulfate, and 20 μg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1X SSC at about 65°C.

Also contemplated are nucleic acid molecules that hybridize to the polynucleotides of the present invention at lower stringency hybridization conditions. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages of formamide result in lowered stringency); salt conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37°C in a solution comprising 6X SSPE (20X SSPE = 3M NaCl; 0.2M NaH_2PO_4; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 μg/ml salmon sperm blocking DNA;

followed by washes at 50°C with 1X SSPE, 0.1% SDS. In addition, to achieve even lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g., 5X SSC).

Note that variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

Of course, a polynucleotide which hybridizes only to polyA+ sequences (such as any 3' terminal polyA+ tract of a cDNA shown in the sequence listing), or to a complementary stretch of T (or U) residues, would not be included in the definition of "polynucleotide," since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

A polynucleotide of the present invention can be composed of any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, a polynucleotide can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. A polynucleotide may also contain one or more modified bases or DNA or RNA backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications can be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically, or metabolically modified forms.

The polypeptide of the present invention can be composed of amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as posttranslational processing, or by chemical modification

techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched, for example, as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched, and branched cyclic polypeptides may result from posttranslation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formulation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. (See, e.g., T. E. Creighton, Ed., *Proteins – Structure And Molecular Properties*, 2nd Ed., W. H. Freeman and Company, New York (1993); B. C. Johnson, Ed., *Posttranslational Covalent Modification Of Proteins*, Academic Press, New York, pgs. 1-12 (1983); Seifter et al., *Meth. Enzymol.* 182:626-646 (1990); Rattan et al., *Ann. N.Y. Acad. Sci.* 663:48-62 (1992)).

"A polypeptide having biological activity" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose-dependency. In the case where dose dependency does exist, it need not be identical to that of the polypeptide, but rather substantially similar to the dose-dependence in a given activity as compared to the polypeptide of the present invention (i.e., the candidate polypeptide will exhibit greater activity or not more than about 25-fold less and, preferably, not more than about ten-fold less activity and, most preferably, not more than about three-fold less activity relative to the polypeptide of the present invention).

The translated amino acid sequence, beginning with the methionine, is identified although other reading frames can also be easily translated using known molecular biology techniques. The polypeptides produced by the translation of these alternative open reading frames are specifically contemplated by the present invention.

SEQ ID NOs:1-23 and the translations of SEQ ID NOs: 1-23 are sufficiently accurate and otherwise suitable for a variety of uses well known in the art and described further below. These nucleic acid molecules will also hybridize to nucleic acid molecules in biological samples, thereby enabling a variety of forensic and diagnostic methods of the invention. Similarly, polypeptides identified from the translations of SEQ ID NOs:1-23 may be used to generate antibodies which bind specifically to the secreted proteins encoded by the cDNA clones identified.

Nevertheless, DNA sequences generated by sequencing reactions can contain sequencing errors. The errors exist as misidentified nucleotides, or as insertions or deletions of nucleotides in the generated DNA sequence. The erroneously inserted or deleted nucleotides cause frame shifts in the reading frames of the predicted amino acid sequence. In these cases, the predicted amino acid sequence diverges from the actual amino acid sequence, even though the generated DNA sequence may be greater than 99.9% identical to the actual DNA sequence (for example, one base insertion or deletion in an open reading frame of over 1,000 bases).

The present invention also relates to the genes corresponding to polynucleotides identified in SEQ ID NOs:1-23, and translations of polynucleotides identified in SEQ ID NOs:1-23. The corresponding gene can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include preparing probes or primers from the disclosed sequence and identifying or amplifying the corresponding gene from appropriate sources of genomic material.

Also provided in the present invention are species homologues. Species homologues may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for the desired homologue.

The polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

The polypeptides may be in the form of the secreted protein, including the mature form, or may be a part of a larger protein, such as a fusion protein (see below). It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification (such as multiple histidine residues), or an additional sequence for stability during recombinant production.

The polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced version of a polypeptide, including the secreted polypeptide, can be substantially purified by the one-step method described in Smith & Johnson, *Gene*, 67:31-40 (1988). Polypeptides of the invention also can be purified from natural or recombinant sources using antibodies of the invention raised against the secreted protein in methods which are well known in the art.

__Signal Sequences__

Methods for predicting whether a protein has a signal sequence, as well as the cleavage point for that sequence, are available. For instance, the method of McGeoch uses the information from a short N-terminal charged region and a subsequent uncharged region of the complete (uncleaved) protein (*Virus Res.*, 3:271-286 (1985)). The method of von Heinje uses the information from the residues surrounding the cleavage site, typically residues -13 to +2, where +1 indicates the amino terminus of the secreted protein (*Nucleic Acids Res.* 14:4683-4690 (1986)). Therefore, from a deduced amino acid sequence, a signal sequence and mature sequence can be identified.

In the present case, the deduced amino acid sequence of the secreted polypeptide was analyzed by a computer program called Signal P (Nielsen et al., *Protein Engineering*, 10:1-6 (1997), which predicts the cellular location of a protein based on the amino acid sequence. As part of this computational prediction of localization, the methods of McGeoch and von Heinje are incorporated.

As one of ordinary skill would appreciate, however, cleavage sites sometimes vary from organism to organism and cannot be predicted with absolute certainty. Accordingly, the present invention provides secreted polypeptides having a sequence corresponding to the translations of SEQ. ID NO:1-23 which have an N-terminus beginning within 5 residues (i.e., + or - 5 residues) of the predicted cleavage point. Similarly, it is also recognized that in some cases, cleavage of the signal sequence from a secreted protein is not entirely uniform, resulting in more than one secreted species. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Moreover, the signal sequence identified by the above analysis may not necessarily predict the naturally occurring signal sequence. For example, the naturally occurring signal sequence may be further upstream from the predicted signal sequence. However, it is likely that the predicted signal sequence will be capable of directing the secreted protein to the ER. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Polynucleotide and Polypeptide Variants

"Variant" refers to a polynucleotide or polypeptide differing from the polynucleotide or polypeptide of the present invention, but retaining essential properties thereof. In general, variants have close similarity overall and are identical in many regions to the polynucleotide or polypeptide of the present invention.

"Identity" per se has an art-recognized meaning and can be calculated using published techniques. (See, e.g., Lesk, Ed., *Computational Molecular Biology*, Oxford University Press, New York, (1988); Smith, Ed., *Biocomputing: Informatics And Genome Projects*, Academic Press, New York, (1993); Griffin, and Griffin, Eds., *Computer Analysis Of Sequence Data, Part I*, Humana Press, New Jersey, (1994); von Heinje, *Sequence Analysis In Molecular Biology*, Academic Press, (1987); and Gribskov and Devereux, Eds., *Sequence Analysis Primer*, M Stockton Press, New York, (1991)). While there exists a number of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans (Carillo et al., *SIAM J Applied Math.*, 48:1073 (1988)). Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to, those disclosed in "Guide to Huge Computers," Martin J. Bishop, Ed., Academic Press, San Diego, (1994) and Carillo et al., (1988), *Supra*. Methods for aligning polynucleotides or polypeptides are codified in computer programs, including the GCG program package (Devereux et al., *Nuc. Acids Res.* 12:387 (1984)), BLASTP, BLASTN, FASTA (Atschul et al., *J. Molec. Biol.* 215:403 (1990)), and Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711) which uses the local homology algorithm of Smith and Waterman (*Adv. in App. Math.*, 2:482-489 (1981)).

When using any of the sequence alignment programs to determine whether a particular sequence is, for instance, 95% identical to a reference sequence, the parameters are set such that the percentage of identity is calculated over the full length of the reference polynucleotide and

that gaps in identity of up to 5% of the total number of nucleotides in the reference polynucleotide are allowed.

A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (*Comp. App. Biosci.* 6:237-245 (1990)). The term "sequence" includes nucleotide and amino acid sequences. In a sequence alignment the query and subject sequences are either both nucleotide sequences or both amino acid sequences. The result of said global sequence alignment is presented in terms of percent identity. Preferred parameters used in a FASTDB search of a DNA sequence to calculate percent identity are: Matrix=Unitary, k-tuple=4, Mismatch Penalty=l, Joining Penalty=30, Randomization Group Length=0, and Cutoff Score=l, Gap Penalty=5, Gap Size Penalty 0.05, and Window Size=500 or query sequence length in nucleotide bases, whichever is shorter. Preferred parameters employed to calculate percent identity and similarity of an amino acid alignment are: Matrix=PAM 150, k-tuple=2, Mismatch Penalty= 1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=l, Gap Penalty=5, Gap Size Penalty=0.05, and Window Size=500 or query sequence length in amino acid residues, whichever is shorter.

As an illustration, a polynucleotide having a nucleotide sequence of at least 95% "identity" to a sequence identified in SEQ ID NOs:1-23 means that the polynucleotide is identical to a sequence identified in SEQ ID NOs:1-23 or the cDNA except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the total length (not just within a given 100 nucleotide stretch). In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to polynucleotides identified in SEQ ID NOs:1-23, up to 5% of the nucleotides in the sequence identified in SEQ ID NOs:1-23 or the cDNA can be deleted, inserted, or substituted with other nucleotides. These changes may occur anywhere throughout the polynucleotide.

Further embodiments of the present invention include polynucleotides having at least 80% identity, more preferably at least 90% identity, and most preferably at least 95%, 96%, 97%, 98% or 99% identity to a sequence identified in SEQ ID NOs:1-23. Of course, due to the degeneracy of the genetic code, one of ordinary skill in the art will immediately recognize that a large number of the polynucleotides having at least 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity will encode a polypeptide identical to an amino acid sequence contained in the translations of polynucleotides identified in SEQ ID NOs:1-23.

Similarly, by a polypeptide having an amino acid sequence having at least, for example, 95% "identity" to a reference polypeptide, is intended that the amino acid sequence of the polypeptide is identical to the reference polypeptide except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the total length of the reference polypeptide. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino acid sequence, up to 5% of the amino acid residues in the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

Further embodiments of the present invention include polypeptides having at least 80% identity, more preferably at least 85% identity, more preferably at least 90% identity, and most preferably at least 95%, 96%, 97%, 98% or 99% identity to an amino acid sequence contained in translations of polynucleotides identified in SEQ ID NOs:1-23. Preferably, the above polypeptides should exhibit at least one biological activity of the protein.

In a preferred embodiment, polypeptides of the present invention include polypeptides having at least 90% similarity, more preferably at least 95% similarity, and still more preferably at least 96%, 97%, 98%, or 99% similarity to an amino acid sequence contained in translations of polynucleotides identified in SEQ ID NOs:1-23.

The variants may contain alterations in the coding regions, non-coding regions, or both. Especially preferred are polynucleotide variants containing alterations which produce silent substitutions, additions, or deletions, but do not alter the properties or activities of the encoded polypeptide. Nucleotide variants produced by silent substitutions due to the degeneracy of the genetic code are preferred. Moreover, variants in which 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination are also preferred. Polynucleotide variants can be produced for a variety of reasons. For instance, a polynucleotide variant may be produced to optimize codon expression for a particular host (i.e., codons in the human mRNA may be changed to those preferred by a bacterial host, such as *E. coli*).

Naturally occurring variants are called "allelic variants," and refer to one of several alternate forms of a gene occupying a given locus on a chromosome of an organism (Lewin, B.,

Ed., *Genes II*, John Wiley & Sons, New York (1985)). These allelic variants can vary at either the polynucleotide and/or polypeptide level. Alternatively, non-naturally occurring variants may be produced by mutagenesis techniques or by direct synthesis.

Using known methods of protein engineering and recombinant DNA technology, variants may be generated to improve or alter the characteristics of the polypeptides of the present invention. For instance, one or more amino acids can be deleted from the N-terminus or C-terminus of the secreted protein without substantial loss of biological function. Ron et al. reported variant KGF proteins having heparin binding activity even after deleting 3, 8, or 27 amino-terminal amino acid residues. (*J. Biol. Chem.* 268: 2984-2988 (1993)). Similarly, interferon gamma exhibited up to ten times higher activity after deleting 8-10 amino acid residues from the carboxy terminus of this protein (Dobeli et al., *J. Biotechnology,* 7:199-216 (1988)).

Moreover, ample evidence demonstrates that variants often retain a biological activity similar to that of the naturally occurring protein. For example, Gayle et al. conducted extensive mutational analysis of human cytokine IL-la (*J. Biol. Chem.,* 268:22105-22111 (1993)). These investigators used random mutagenesis to generate over 3,500 individual IL-la mutants that averaged 2.5 amino acid changes per variant over the entire length of the molecule. Multiple mutations were examined at every possible amino acid position. The investigators concluded that "[m]ost of the molecule could be altered with little effect on either [binding or biological activity]." (See, Gayle et al. (1993), Abstract). In fact, only 23 unique amino acid sequences, out of more than 3,500 nucleotide sequences examined, produced a protein that differed significantly in activity from the wild-type sequence.

Furthermore, even if deleting one or more amino acids from the N-terminus or C-terminus of a polypeptide results in modification or loss of one or more biological functions, other biological activities may still be retained. For example, the ability of a deletion variant to induce and/or to bind antibodies which recognize the secreted form will likely be retained when less than the majority of the residues of the secreted form are removed from the N-terminus or C-terminus. Whether a particular polypeptide lacking N- or C-terminal residues of a protein retains such immunogenic activities can readily be determined by routine methods described herein and otherwise known in the art.

Thus, the invention further includes polypeptide variants which show substantial biological activity. Such variants include deletions, insertions, inversions, repeats, and substitutions selected according to general rules known in the art so as have little effect on

activity. For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie, et al., *Science* 247:1306-1310 (1990), wherein the authors indicate that there are two main strategies for studying the tolerance of an amino acid sequence to change.

The first strategy exploits the tolerance of amino acid substitutions by natural selection during the process of evolution. By comparing amino acid sequences in different species, the amino acid positions which have been conserved between species can be identified. These conserved amino acids are likely important for protein function. In contrast, the amino acid positions in which substitutions have been tolerated by natural selection indicate positions which are not critical for protein function. Thus, positions tolerating amino acid substitution may be modified while still maintaining biological activity of the protein.

The second strategy uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene to identify regions critical for protein function. For example, site-directed mutagenesis or alanine-scanning mutagenesis (the introduction of single alanine mutations at every residue in the molecule) can be used (Cunningham et al., *Science* 244:1081-1085 (1989)). The resulting mutant molecules can then be tested for biological activity.

According to Bowie et al. (1990), these two strategies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at certain amino acid positions in the protein. For example, most buried or interior (within the tertiary structure of the protein) amino acid residues require nonpolar side chains, whereas few features of surface or exterior side chains are generally conserved. Moreover, tolerated conservative amino acid substitutions involve replacement of the aliphatic or hydrophobic amino acids Ala, Val, Leu and Ile; replacement of the hydroxyl residues Ser and Thr; replacement of the acidic residues Asp and Glu; replacement of the amide residues Asn and Gln; replacement of the basic residues Lys, Arg, and His; replacement of the aromatic residues Phe, Tyr, and Trp; and replacement of the small-sized amino acids Ala, Ser, Thr, Met, and Gly.

Besides conservative amino acid substitution, variants of the present invention include: (i) substitutions with one or more of the non-conserved amino acid residues, where the substituted amino acid residues may or may not be one encoded by the genetic code; (ii) substitution with one or more of amino acid residues having a substituent group; (iii) fusion of the mature polypeptide with another compound, such as a compound to increase the stability and/or solubility of the polypeptide (e.g., polyethylene glycol); or (iv) fusion of the polypeptide

with additional amino acids, such as an IgG Fc fusion region peptide, a leader or secretory sequence, or a sequence facilitating purification. Such variant polypeptides are deemed to be within the scope of those skilled in the art from the teachings herein.

For example, polypeptide variants containing amino acid substitutions of charged amino acids with other charged or neutral amino acids may produce proteins with improved characteristics, such as decreased aggregation. As known, aggregation of pharmaceutical formulations both reduces activity and increases clearance due to the aggregate's immunogenic activity (see, e.g., Pinckard et al., *Clin. Exp. Immunol.* 2:331-340 (1967); Robbins et al., *Diabetes* 36: 838-845 (1987); Cleland et al., *Crit. Rev. Therap. Drug Carrier Sys.* 10:307-377 (1993)).

Polynucleotide and Polypeptide Fragments

In the present invention, a "polynucleotide fragment" refers to a short polynucleotide having a nucleic acid sequence contained in that shown in polynucleotides identified in SEQ ID NOs:1-23. The short nucleotide fragments are preferably at least about 15 nucleotides (nt), and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length. A fragment "at least 20 nt in length," for example, is intended to include 20 or more contiguous bases from the cDNA sequence contained in that shown in polynucleotides identified in SEQ ID NOs:1-23. These nucleotide fragments are useful as diagnostic probes and primers as discussed herein. Of course, larger fragments (e.g., 40, 50, 150, and greater than 150 nucleotides) are preferred.

Moreover, representative examples of polynucleotide fragments of the invention, include, for example, fragments having a sequence from about nucleotide number 1-40, 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, to the end of polynucleotides identified in SEQ ID NOs:1-23. In this context "about" includes the particularly recited ranges, larger or smaller by several nucleotides (i.e., 5, 4, 3, 2, or 1 nt), at either terminus or at both termini. Preferably, these fragments encode a polypeptide which has biological activity.

In the present invention, a "polypeptide fragment" refers to a short amino acid sequence contained in the translations of polynucleotides identified in SEQ ID NOs:1-23. Protein fragments may be "free-standing," or comprised within a larger polypeptide of which the fragment forms a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention, include, for example, fragments from about amino acid number 1-20, 21-40, 41-60, or 61 to the end of the coding region. Moreover,

polypeptide fragments can be about 20, 30, 40, 50 or 60, amino acids in length. In this context "about" includes the particularly recited ranges, larger or smaller by several amino acids (i.e., 5, 4, 3, 2, or 1), at either extreme or at both extremes.

Preferred polypeptide fragments include the secreted protein as well as the mature form. Further preferred polypeptide fragments include the secreted protein or the mature form having a continuous series of deleted residues from the amino or the carboxy terminus, or both. For example, any number of amino acids ranging from 1-60, can be deleted from the amino terminus of either the secreted polypeptide or the mature form. Similarly, any number of amino acids, ranging from 1-30, can be deleted from the carboxy terminus of the secreted protein or mature form. Furthermore, any combination of the above amino and carboxy terminus deletions are preferred. Similarly, polynucleotide fragments encoding these polypeptide fragments are also preferred.

Also preferred are polypeptide and polynucleotide fragments characterized by structural or functional domains, such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions. Polypeptide fragments of the translations of polynucleotides identified in SEQ ID NOs:1-23 falling within conserved domains are specifically contemplated by the present invention. Moreover, polynucleotide fragments encoding these domains are also contemplated.

Other preferred fragments are biologically active fragments. Biologically active fragments are those exhibiting activity similar, but not necessarily identical, to an activity of the polypeptide of the present invention. The biological activity of the fragments may include an improved desired activity, or a decreased undesirable activity.

Epitopes & Antibodies

In the present invention, "epitopes" refer to polypeptide fragments having antigenic or immunogenic activity in an animal, especially in a human. A preferred embodiment of the present invention relates to a polypeptide fragment comprising an epitope, as well as the polynucleotide encoding this fragment. A region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." In contrast, an "immunogenic epitope" is defined as a part of a protein that elicits an antibody response. (See, e.g. Geysen et al., *Proc. Natl. Acad. Sci. USA,* 81:3998-4002 (1983)).

Fragments which function as epitopes may be produced by any conventional means. (See, e.g., Houghten, R. A., *Proc. Natl. Acad. Sci. USA*, 82:5131-5135 (1985) further described in U.S. Patent No. 4,631,211).

In the present invention, antigenic epitopes preferably contain a sequence of at least seven, more preferably at least nine, and most preferably between about 15 to about 30 amino acids. Antigenic epitopes are useful to raise antibodies, including monoclonal antibodies, that specifically bind the epitope. (See, e.g., Wilson et al., *Cell,* 37:767-778 (1984); Sutcliffe et al., *Science,* 219:660-666 (1983)).

Similarly, immunogenic epitopes can be used to induce antibodies according to methods well known in the art. (See, e.g., Sutcliffe et al., (1983) *Supra*; Wilson et al., (1984) *Supra*; Chow et al., *Proc. Natl. Acad. Sci. USA,* 82:910-914; and Bittle et al., *J. Gen. Virol.*, 66:2347-2354 (1985)). A preferred immunogenic epitope includes the secreted protein. The immunogenic epitope may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse). Alternatively, the immunogenic epitope may be prescribed without a carrier if the sequence is of sufficient length (at least about 25 amino acids). However, immunogenic epitopes comprising as few as 8 to 10 amino acids have been shown to be sufficient to raise antibodies capable of binding to, at the very least, linear epitopes in a denatured polypeptide (e.g., in Western blotting.)

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab')2 fragments) which are capable of specifically binding to protein. Fab and F(ab')2 fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding than an intact antibody. (Wahl et al., *J. Nucl. Med.* 24:316-325 (1983)). Thus, these fragments are preferred, as well as the products of a Fab or other immunoglobulin expression library. Moreover, antibodies of the present invention include chimeric, single chain, human, and humanized antibodies.

Additional embodiments include chimeric antibodies, e.g., humanized versions of murine monoclonal antibodies. Such humanized antibodies may be prepared by known techniques, and offer the advantage of reduced immunogenicity when the antibodies are administered to humans. In one embodiment, a humanized monoclonal antibody comprises the variable region of a murine antibody (or just the antigen binding site thereof) and a constant

region derived from a human antibody. Alternatively, a humanized antibody fragment may comprise the antigen binding site of a murine monoclonal antibody and a variable region fragment (lacking the antigen-binding site) derived from a human antibody. Procedures for the production of chimeric and further engineered monoclonal antibodies include those described in Riechmann et al. (*Nature,* 332:323, 1988), Liu et al. (*PNAS,* 84:3439, 1987), Larrick et al. (*Bio/Technology,* 7:934, 1989), and Winter and Harris (*TIPS,* 14:139, May, 1993).

One method for producing an antibody comprises immunizing a non-human animal, such as a transgenic mouse, with a polypeptide translated from a nucleotide sequence chosen from SEQ ID NOs:1-23, whereby antibodies directed against the polypeptide translated from a nucleotide sequence chosen from SEQ ID NOs:1-23 are generated in said animal. Procedures have been developed for generating human antibodies in non-human animals. The antibodies may be partially human, or preferably completely human. Non-human animals (such as transgenic mice) into which genetic material encoding one or more human immunoglobulin chains has been introduced may be employed. Such transgenic mice may be genetically altered in a variety of ways. The genetic manipulation may result in human immunoglobulin polypeptide chains replacing endogenous immunoglobulin chains in at least some (preferably virtually all) antibodies produced by the animal upon immunization. Antibodies produced by immunizing transgenic animals with a polypeptide translated from a nucleotide sequence chosen from SEQ ID NOs:1-23 are provided herein.

Mice in which one or more endogenous immunoglobulin genes are inactivated by various means have been prepared. Human immunoglobulin genes have been introduced into the mice to replace the inactivated mouse genes. Antibodies produced in the animals incorporate human immunoglobulin polypeptide chains encoded by the human genetic material introduced into the animal. Examples of techniques for production and use of such transgenic animals are described in U.S. Patent Nos. 5,814,318, 5,569,825, and 5,545,806, which are incorporated by reference herein.

Monoclonal antibodies may be produced by conventional procedures, e.g., by immortalizing spleen cells harvested from the transgenic animal after completion of the immunization schedule. The spleen cells may be fused with myeloma cells to produce hybridomas, by conventional procedures.

A method for producing a hybridoma cell line comprises immunizing such a transgenic animal with a immunogen comprising at least seven contiguous amino acid residues of a polypeptide translated from a nucleotide sequence chosen from SEQ ID NOs:1-23; harvesting spleen cells from the immunized animal; fusing the harvested spleen cells to a myeloma cell line, thereby generating hybridoma cells; and identifying a hybridoma cell line that produces a monoclonal antibody that binds a polypeptide translated from a nucleotide sequence chosen from SEQ ID NOs:1-23. Such hybridoma cell lines, and monoclonal antibodies produced therefrom, are encompassed by the present invention. Monoclonal antibodies secreted by the hybridoma cell line are purified by conventional techniques.

Antibodies may be employed in an *in vitro* procedure, or administered *in vivo* to inhibit biological activity induced by a polypeptide translated from a nucleotide sequence chosen from SEQ ID NOs:1-23. Disorders caused or exacerbated (directly or indirectly) by the interaction of such polypeptides of the present invention with cell surface receptors thus may be treated. A therapeutic method involves *in vivo* administration of a blocking antibody to a mammal in an amount effective for reducing a biological activity induced by a polypeptide translated from a nucleotide sequence chosen from SEQ ID NOs:1-23.

Also provided herein are conjugates comprising a detectable (e.g., diagnostic) or therapeutic agent, attached to an antibody directed against a polypeptide translated from a nucleotide sequence chosen from SEQ ID NOs:1-23. Examples of such agents are well known, and include but are not limited to diagnostic radionuclides, therapeutic radionuclides, and cytotoxic drugs. The conjugates find use in *in vitro* or *in vivo* procedures.

Fusion Proteins

Any polypeptide of the present invention can be used to generate fusion proteins. For example, the polypeptide of the present invention, when fused to a second protein, can be used as an antigenic tag. Antibodies raised against the polypeptide of the present invention can be used to indirectly detect the second protein by binding to the polypeptide. Moreover, because secreted proteins target cellular locations based on trafficking signals, the polypeptides of the present invention can be used as targeting molecules once fused to other proteins.

Examples of domains that can be fused to polypeptides of the present invention include not only heterologous signal sequences, but also other heterologous functional regions. The fusion does not necessarily need to be direct, but may occur through linker sequences.

Moreover, fusion proteins may also be engineered to improve characteristics of the polypeptide of the present invention. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence during purification from the host cell or subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to facilitate handling of polypeptides are familiar and routine techniques in the art.

In addition, polypeptides of the present invention, including fragments and specifically epitopes, can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins facilitate purification and show an increased half-life in vivo. One reported example describes chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins. (EP A 394,827; Traunecker et al., *Nature*, 331:84-86 (1988)). Fusion proteins having disulfide-linked dimeric structures (due to the IgG) can also be more efficient in binding and neutralizing other molecules, than the monomeric secreted protein or protein fragment alone. (Fountoulakis et al., *J. Biochem.*, 270:3958-3964 (1995)).

Similarly, EP A 0 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is beneficial in therapy and diagnosis, and thus can result in, for example, improved pharmacokinetic properties (see, e.g., EP A 0 232 262). Alternatively, deleting the Fc part after the fusion protein has been expressed, detected, and purified, would be desired. For example, the Fc portion may hinder therapy and diagnosis if the fusion protein is used as an antigen for immunizations. In drug discovery, for example, human proteins, such as hIL-5, have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5. (See, Bennett et al., *J. Mol. Recognition* 8:52-58 (1995); Johanson et al., *J. Biol. Chem.*, 270:9459-9471 (1995)).

Moreover, the polypeptides of the present invention can be fused to marker sequences, such as a peptide which facilitates purification of the fused polypeptide. In preferred embodiments, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., Chatsworth, CA), among others, many of which are

commercially available. As described in Gentz et al., for instance, hexa-histidine provides for convenient purification of the fusion protein (*Proc. Natl. Acad. Sci. USA* 86:821-824 (1989)). Another peptide tag useful for purification, the "HA" tag, corresponds to an epitope derived from the influenza hemagglutinin protein. (Wilson et al., *Cell* 37:767 (1984)).

Thus, any of these above fusions can be engineered using the polynucleotides or the polypeptides of the present invention.

<u>Vectors, Host Cells, and Protein Production</u>

The present invention also relates to vectors containing the polynucleotide of the present invention, host cells, and the production of polypeptides by recombinant techniques. The vector may be, for example, a phage, plasmid, viral, or retroviral vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged in vitro using an appropriate packaging cell line and then transduced into host cells.

The polynucleotide insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the *E. coli* lac, trp, phoA and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the transcripts expressed by the constructs will preferably include a translation initiating codon at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture and tetracycline, kanamycin or ampicillin resistance genes for culturing in *E. coli* and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as *E. coli*, *Streptomyces* and *Salmonella typhimurium* cells; fungal cells, such as yeast cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells;

animal cells such as CHO, COS, 293, and Bowes melanoma cells, and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from QIAGEN, Inc.; pBluescript vectors, Phagescript vectors, pNH8A, PNH16a, PNH18A, pNH46A, available from Stratagene Cloning Systems, Inc.; and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia Biotech, Inc. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXTl and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or other methods. Such methods are described in many standard laboratory manuals, such as Davis et al., *Basic Methods In Molecular Biology* (1986). It is specifically contemplated that the polypeptides of the present invention may, in fact, be expressed by a host cell lacking a recombinant vector.

A polypeptide of this invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification.

Polypeptides of the present invention, and preferably the secreted form, can also be recovered from products purified from natural sources, including bodily fluids, tissues and cells, whether directly isolated or cultured; products of chemical synthetic procedures; and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect, and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes. Thus, it is well known in the art that the N-terminal methionine encoded by the translation initiation codon generally is removed with high efficiency from any protein after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some proteins, this prokaryotic removal process

is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

Uses of the Polynucleotides

Each of the polynucleotides identified herein can be used in numerous ways as reagents. The following description should be considered exemplary and utilizes known techniques.

The polynucleotides of the present invention are useful for chromosome identification. There exists an ongoing need to identify new chromosome markers, since few chromosome marking reagents based on actual sequence data (repeat polymorphisms) are presently available. Each polynucleotide of the present invention can be used as a chromosome marker.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the sequences shown in polynucleotides identified in SEQ ID NOs:1-23. Primers can be selected using computer analysis so that primers do not span more than one predicted exon in the genomic DNA. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the polynucleotides identified in SEQ ID NOs:1-23 will yield an amplified fragment.

Similarly, somatic hybrids provide a rapid method of PCR mapping the polynucleotides to particular chromosomes. Three or more clones can be assigned per day using a single thermal cycler. Moreover, sublocalization of the polynucleotides can be achieved with panels of specific chromosome fragments. Other gene-mapping strategies that can be used include *in situ* hybridization, prescreening with labeled flow-sorted chromosomes, and preselection by hybridization to construct chromosome specific-cDNA libraries.

Precise chromosomal location of the polynucleotides can also be achieved using fluorescence in situ hybridization (FISH) of a metaphase chromosomal spread. This technique uses polynucleotides as short as 500 or 600 bases; however, polynucleotides of 2,000-4,000 bp are preferred. For a review of this technique, see Verma et al., *Human Chromosomes: a Manual of Basic Techniques*, Pergamon Press, New York (1988).

For chromosome mapping, the polynucleotides can be used individually (to mark a single chromosome or a single site on that chromosome) or in panels (for marking multiple sites and/or multiple chromosomes). Preferred polynucleotides correspond to the noncoding regions

of the cDNAs because the coding sequences are more likely conserved within gene families, thus increasing the chance of cross-hybridization during chromosomal mapping.

Once a polynucleotide has been mapped to a precise chromosomal location, the physical position of the polynucleotide can be used in linkage analysis. Linkage analysis establishes coinheritance between a chromosomal location and presentation of a particular disease. Disease mapping data are found, for example, in V. McKusick, *Mendelian Inheritance in Man* (available on line through Johns Hopkins University Welch Medical Library)). Assuming one megabase mapping resolution and one gene per 20 kb, a cDNA precisely localized to a chromosomal region associated with the disease could be one of 50-500 potential causative genes.

Thus, once coinheritance is established, differences in the polynucleotide and the corresponding gene between affected and unaffected individuals can be examined. The polynucleotides of polynucleotides identified in SEQ ID NOs:1-23 can be used for this analysis of individuals.

First, visible structural alterations in the chromosomes, such as deletions or translocations, are examined in chromosome spreads or by PCR. If no structural alterations exist, the presence of point mutations are ascertained. Mutations observed in some or all affected individuals, but not in normal individuals, indicates that the mutation may cause the disease. However, complete sequencing of the polypeptide and the corresponding gene from several normal individuals is required to distinguish the mutation from a polymorphism. If a new polymorphism is identified, this polymorphic polypeptide can be used for further linkage analysis.

Furthermore, increased or decreased expression of the gene in affected individuals as compared to unaffected individuals can be assessed using polynucleotides of the present invention. Any of these alterations (altered expression, chromosomal rearrangement, or mutation) can be used as a diagnostic or prognostic marker.

In addition to the foregoing, a polynucleotide can be used to control gene expression through triple helix formation or antisense DNA or RNA. Both methods rely on binding of the polynucleotide to DNA or RNA. For these techniques, preferred polynucleotides are usually 20 to 40 bases in length and complementary to either the region of the gene involved in transcription (see, Lee et al., *Nuc. Acids Res.*, 6:3073 (1979); Cooney et al., *Science,* 241:456 (1988); and Dervan et al., *Science,* 251:1360 (1991) for discussion of triple helix formation) or

to the mRNA itself (see, Okano, *J. Neurochem.*, 56:560 (1991); and *Oligodeoxy-nucleotides as Antisense Inhibitors of Gene Expression*, CRC Press, Boca Raton, FL (1988) (for a discussion of antisense technique). Triple helix formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques are effective in model systems, and the information disclosed herein can be used to design antisense or triple helix polynucleotides in an effort to treat disease.

Polynucleotides of the present invention are also useful in gene therapy. One goal of gene therapy is to insert a normal gene into an organism having a defective gene, in an effort to correct the genetic defect. The polynucleotides disclosed in the present invention offer a means of targeting such genetic defects in a highly accurate manner. Another goal is to insert a new gene that was not present in the host genome, thereby producing a new trait in the host cell.

The polynucleotides are also useful for identifying individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identifying personnel. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The polynucleotides of the present invention can be used as additional DNA markers for RFLP.

The polynucleotides of the present invention can also be used as an alternative to RFLP, by determining the actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can be used to prepare PCR primers for amplifying and isolating such selected DNA, which can then be sequenced. Using this technique, individuals can be identified because each individual will have a unique set of DNA sequences. Once an unique ID database is established for an individual, positive identification of that individual, living or dead, can be made from extremely small tissue samples.

Forensic biology also benefits from using DNA-based identification techniques as disclosed herein. DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, semen, etc., can be amplified using PCR. In one prior art technique, gene sequences amplified from polymorphic loci, such as DQa class II HLA gene, are used in forensic biology to identify individuals. (Erlich, Ed., *PCR Technology*, M. Stockton Press (1989)). Once these specific polymorphic loci are amplified, they are

digested with one or more restriction enzymes, yielding an identifying set of bands on a Southern blot probed with DNA corresponding to the DQa class H HLA gene. Similarly, polynucleotides of the present invention can be used as polymorphic markers for forensic purposes.

There is also a need for reagents capable of identifying the source of a particular tissue. Such need arises, for example, in forensics when presented with tissue of unknown origin. Appropriate reagents can comprise, for example, DNA probes or primers specific to particular tissue prepared from the sequences of the present invention. Panels of such reagents can identify tissue by species and/or by organ type. In a similar fashion, these reagents can be used to screen tissue cultures for contamination.

In the very least, the polynucleotides of the present invention can be used as molecular weight markers on Southern gels; as diagnostic probes for the presence of a specific mRNA in a particular cell type; as a probe to "subtract-out" known sequences in the process of discovering novel polynucleotides; for selecting and making oligomers for attachment to a "gene chip" or other support; to raise anti-DNA antibodies using DNA immunization techniques; and as an antigen to elicit an immune response.

Uses of the Polypeptides

Each of the polypeptides identified herein can be used in numerous ways. The following description should be considered exemplary and utilizes known techniques.

A polypeptide of the present invention can be used to assay protein levels in a biological sample using antibody-based techniques. For example, protein expression in tissues can be studied with classical immunohistological methods. (Jalkanen, et al., *J. Cell. Biol.*, 101:976-985 (1985); Jalkanen, et al., *J. Cell Biol.*, 105:3087-3096 (1987)). Other antibody-based methods useful for detecting protein gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). Suitable antibody assay labels are known in the art and include enzyme labels, such as, glucose oxidase, and radioisotopes, such as iodine (^{125}I, ^{121}I), carbon (^{14}C), sulfur (^{35}S), tritium (^{3}H), indium (^{112}In), and technetium (^{99m}Tc), and fluorescent labels, such as fluorescein and rhodamine, and biotin.

In addition to assaying secreted protein levels in a biological sample, proteins can also be detected in vivo by imaging. Antibody labels or markers for *in vivo* imaging of protein include those detectable by X-radiography, nuclear magnetic resonance (NMR) or ESR. For X-radiography, suitable labels include radioisotopes such as barium or cesium, which emit

detectable radiation but are not overtly harmful to the subject. Suitable markers for NMR and ESR include those with a detectable characteristic spin, such as deuterium, which may be incorporated into the antibody by labeling of nutrients for the relevant hybridoma.

A protein-specific antibody or antibody fragment which has been labeled with an appropriate detectable imaging moiety, such as a radioisotope (e.g., ^{131}I, ^{112}In, ^{99m}Tc), a radio-opaque substance, or a material detectable by NMR, is introduced (e.g., parenterally, subcutaneously, or intraperitoneally) into the mammal. It will be understood in the art that the size of the subject and the imaging system used will determine the quantity of imaging moiety needed to produce diagnostic images. In the case of a radioisotope moiety, the quantity of radioactivity necessary for a human subject will normally range from about 5 to 20 millicuries of ^{99m}Tc. The labeled antibody or antibody fragment will then preferentially accumulate at the location of cells which contain the specific protein. In vivo tumor imaging is described in Burchiel et al., "Immunopharmacokinetics of Radiolabeled Antibodies and Their Fragments" (Chapter 13 in *Tumor Imaging: The Radiochemical Detection of Cancer*, Burchiel and Rhodes, Eds., Masson Publishing Inc. (1982).)

Thus, the invention provides a diagnostic method of a disorder, which involves (a) assaying the expression of a polypeptide of the present invention in cells or body fluid of an individual; and (b) comparing the level of gene expression with a standard gene expression level, whereby an increase or decrease in the assayed polypeptide gene expression level compared to the standard expression level is indicative of a disorder.

Moreover, polypeptides of the present invention can be used to treat disease. For example, patients can be administered a polypeptide of the present invention in an effort to replace absent or decreased levels of the polypeptide (e.g., insulin); to supplement absent or decreased levels of a different polypeptide (e.g., hemoglobin S for hemoglobin B); to inhibit the activity of a polypeptide (e.g., an oncogene); to activate the activity of a polypeptide (e.g., by binding to a receptor); to reduce the activity of a membrane bound receptor by competing with it for free ligand (e.g., soluble TNF receptors used in reducing inflammation); or to bring about a desired response (e.g., blood vessel growth).

Similarly, antibodies directed to a polypeptide of the present invention can also be used to treat disease. For example, administration of an antibody directed to a polypeptide of the present invention can bind and reduce overproduction of the polypeptide. Similarly, administration of an antibody can activate the polypeptide, such as by binding to a polypeptide bound to a membrane (receptor).

At the very least, the polypeptides of the present invention can be used as molecular weight markers on SDS-PAGE gels or on molecular sieve gel filtration columns using methods well-known to those of skill in the art. Polypeptides can also be used to raise antibodies, which in turn are used to measure protein expression from a recombinant cell, as a way of assessing transformation of the host cell. Moreover, the polypeptides of the present invention can be used to test the following biological activities.

Biological Activities

The polynucleotides and polypeptides of the present invention can be used in assays to test for one or more biological activities. If these polynucleotides and polypeptides do exhibit activity in a particular assay, it is likely that these molecules may be involved in the diseases associated with the biological activity. Thus, the polynucleotides and polypeptides could be used to treat the associated disease.

Nervous System Activity

A polypeptide or polynucleotide of the present invention may be useful in treating deficiencies or disorders of the central nervous system or peripheral nervous system by activating or inhibiting the proliferation, differentiation, or mobilization (chemotaxis) of neuroblasts, stem cells or glial cells. Also, a polypeptide or polynucleotide of the present invention may be useful in treating deficiencies or disorders of the central nervous system or peripheral nervous system by activating or inhibiting the mechanisms of synaptic transmission, synthesis, metabolism and inactivation of neural transmitters, neuromodulators and trophic factors, and by activating or inhibiting the expression and incorporation of enzymes, structural proteins, membrane channels and receptors in neurons and glial cells.

The etiology of these deficiencies or disorders may be genetic, somatic (such as cancer or some autoimmune disorder), acquired (e.g., by chemotherapy or toxins), or infectious. Moreover, a polynucleotide or polypeptide of the present invention can be used as a marker or detector of a particular nervous system disease or disorder. The disorder or disease can be any of Alzheimer's disease, Pick's disease, Binswanger's disease, other senile dementia, Parkinson's disease, parkinsonism, obsessive compulsive disorders, epilepsy, encephaolopathy, ischemia, alcohol addiction, drug addiction, schizophrenia, amyotrophic lateral sclerosis, multiple sclerosis, depression, and bipolar manic-depressive disorder. Alternatively, the polypeptide or polynucleotide of the present invention can be used to study circadian variation, aging, or long-term potentiation, the latter affecting the hippocampus. Additionally, particularly with reference to mRNA species occurring in particular structures within the central

nervous system, the polypeptide or polynucleotide of the present invention can be used to study brain regions that are known to be involved in complex behaviors, such as learning and memory, emotion, drug addiction, glutamate neurotoxicity, feeding behavior, olfaction, viral infection, vision, and movement disorders.

Immune Activity

A polypeptide or polynucleotide of the present invention may be useful in treating deficiencies or disorders of the immune system, by activating or inhibiting the proliferation, differentiation, or mobilization (chemotaxis) of immune cells. Immune cells develop through a process called hematopoiesis, producing myeloid (platelets, red blood cells, neutrophils, and macrophages) and lymphoid (B and T lymphocytes) cells from pluripotent stem cells. The etiology of these immune deficiencies or disorders may be genetic, somatic (such as cancer or some autoimmune disorders), acquired (e.g., by chemotherapy or toxins), or infectious. Moreover, a polynucleotide or polypeptide of the present invention can be used as a marker or detector of a particular immune system disease or disorder.

A polynucleotide or polypeptide of the present invention may be useful in treating or detecting deficiencies or disorders of hematopoietic cells. A polypeptide or polynucleotide of the present invention could be used to increase differentiation and proliferation of hematopoietic cells, including the pluripotent stem cells, in an effort to treat those disorders associated with a decrease in certain (or many) types hematopoietic cells. Examples of immunologic deficiency syndromes include, but are not limited to: blood protein disorders (e.g. agammaglobulinemia, dysgammaglobulinemia), ataxia telangiectasia, common variable immunodeficiency, Di George's Syndrome, HIV infection, HTLV-BLV infection, leukocyte adhesion deficiency syndrome, lymphopenia, phagocyte bactericidal dysfunction, severe combined immunodeficiency (SCIDs), Wiskott-Aldrich Disorder, anemia, thrombocytopenia, or hemoglobinuria.

Moreover, a polypeptide or polynucleotide of the present invention could also be used to modulate hemostatic (bleeding cessation) or thrombolytic activity (clot formation). For example, by increasing hemostatic or thrombolytic activity, a polynucleotide or polypeptide of the present invention could be used to treat blood coagulation disorders (e.g., afibrinogenemia, factor deficiencies), blood platelet disorders (e.g. thrombocytopenia), or wounds resulting from trauma, surgery, or other causes. Alternatively, a polynucleotide or polypeptide of the present invention that can decrease hemostatic or thrombolytic activity could be used to inhibit or dissolve clotting. These molecules could be important in the treatment of heart attacks (infarction), strokes, or scarring.

A polynucleotide or polypeptide of the present invention may also be useful in the treatment or detection of autoimmune disorders. Many autoimmune disorders result from inappropriate recognition of self as foreign material by immune cells. This inappropriate recognition results in an immune response leading to the destruction of the host tissue. Therefore, the administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing autoimmune disorders.

Examples of autoimmune disorders that can be treated or detected by the present invention include, but are not limited to: Addison's Disease, hemolytic anemia, antiphospholipid syndrome, rheumatoid arthritis, dermatitis, allergic encephalomyelitis, glomerulonephritis, Goodpasture's Syndrome, Graves' Disease, Multiple Sclerosis, Myasthenia Gravis, Neuritis, Ophthalmia, Bullous Pemphigoid, Pemphigus, Polyendocrinopathies, Purpura, Reiter's Disease, Stiff-Man Syndrome, Autoimmune Thyroiditis, Systemic Lupus Erythematosus, Autoimmune Pulmonary Inflammation, Guillain-Barre Syndrome, insulin dependent diabetes mellitis, and autoimmune inflammatory eye disease.

Similarly, allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems, may also be treated by a polypeptide or polynucleotide of the present invention. Moreover, these molecules can be used to treat anaphylaxis, hypersensitivity to an antigenic molecule, or blood group incompatibility.

A polynucleotide or polypeptide of the present invention may also be used to treat and/or prevent organ rejection or graft-versus-host disease (GVHD). Organ rejection occurs by host immune cell destruction of the transplanted tissue through an immune response. Similarly, an immune response is also involved in GVHD, but, in this case, the foreign transplanted immune cells destroy the host tissues. The administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing organ rejection or GVHD.

Similarly, a polypeptide or polynucleotide of the present invention may also be used to modulate inflammation. For example, the polypeptide or polynucleotide may inhibit the proliferation and differentiation of cells involved in an inflammatory response. These molecules can be used to treat inflammatory conditions, both chronic and acute conditions, including inflammation associated with infection (e.g., septic shock, sepsis, or systemic

inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine induced lung injury, inflammatory bowel disease, Crohn's disease, or resulting from over production of cytokines (e.g., TNF or IL-1.)

Hyperproliferative Disorders

A polypeptide or polynucleotide can be used to treat or detect hyperproliferative disorders, including neoplasms. A polypeptide or polynucleotide of the present invention may inhibit the proliferation of the disorder through direct or indirect interactions. Alternatively, a polypeptide or polynucleotide of the present invention may proliferate other cells which can inhibit the hyperproliferative disorder.

For example, by increasing an immune response, particularly increasing antigenic qualities of the hyperproliferative disorder or by inducing the proliferation, differentiation, or mobilization of T-cells, hyperproliferative disorders can be treated. This immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, decreasing an immune response may also be a method of treating hyperproliferative disorders, such as by administering the polypeptide or polynucleotide as a chemotherapeutic agent.

Examples of hyperproliferative disorders that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but are not limited to neoplasms located in the: abdomen, bone, breast, digestive system, liver, pancreas, peritoneum, endocrine glands (adrenal, parathyroid, pituitary, testicles, ovary, thymus, thyroid), eye, head and neck, nervous system (central and peripheral), lymphatic system, pelvic region, skin, soft tissue, spleen, thoracic region, and urogenital system.

Similarly, other hyperproliferative disorders can also be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of such hyperproliferative disorders include, but are not limited to hypergammaglobulinemia, lymphoproliferative disorders, paraproteinemias, purpura, sarcoidosis, Sezary Syndrome, Waldenstron's Macroglobulinemia, Gaucher's Disease, histiocytosis, and any other hyperproliferative disease, besides neoplasia, located in an organ system listed above.

Infectious Disease

A polypeptide or polynucleotide of the present invention can be used to treat or detect infectious agents. For example, by increasing the immune response, particularly increasing the

proliferation and differentiation of B and/or T cells, infectious diseases may be treated. The immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, the polypeptide or polynucleotide of the present invention may also directly inhibit the infectious agent, without necessarily eliciting an immune response.

Viruses are one example of an infectious agent that can cause disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of viruses, include, but are not limited to the following DNA and RNA viral families Arbovirus, Adenoviridae, Arenaviridae, Arterivirus, Birnaviridae, Bunyaviridae, Caliciviridae, Circoviridae, Coronaviridae, Flaviviridae, Hepadnaviridae (Hepatitis), Herpesviridae (such as, Cytomegalovirus, Herpes Simplex, Herpes Zoster), Mononegavirus (e.g., Paramyxoviridae, Morbillivirus, Rhabdoviridae), Orthomyxoviridae (e.g., Influenza), Papovaviridae, Parvoviridae, Picornaviridae, Poxviridae (such as Smallpox or Vaccinia), Reoviridae (e.g., Rotavirus), Retroviridae (HTLV-I, HTLV-II, Lentivirus), and Togaviridae (e.g., Rubivirus). Viruses falling within these families can cause a variety of diseases or symptoms, including, but not limited to arthritis, bronchiollitis, encephalitis, eye infections (e.g., conjunctivitis, keratitis), chronic fatigue syndrome, hepatitis (A, B, C, E, Chronic Active, Delta), meningitis, opportunistic infections (e.g., AIDS), pneumonia, Burkitt's Lymphoma, chickenpox, hemorrhagic fever, Measles, Mumps, Parainfluenza, Rabies, the common cold, Polio, leukemia, Rubella, sexually transmitted diseases, skin diseases (e.g., Kaposi's, warts), and viremia. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Similarly, bacterial or fungal agents that can cause disease or symptoms and that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but are not limited to, the following Gram-Negative and Gram-positive bacterial families and fungi: Actinomycetales (e.g., Corynebacterium, Mycobacterium, Norcardia), Aspergillosis, Bacillaceae (e.g., Anthrax, Clostridium), Bacteroidaceae, Blastomycosis, Bordetella, Borrelia, Brucellosis, Candidiasis, Campylobacter, Coccidioidomycosis, Cryptococcosis, Dermatocycoses, Enterobacteriaceae (Klebsielia, Salmonella, Serratia, Yersinia), Erysipelothrix, Helicobacter, Legionellosis, Leptospirosis, Listeria, Mycoplasmatales, Neisseriaceae (e.g., Acinetobacter, Gonorrhea, Menigococcal), Pasteurellacea infections (e.g., Actinobacillus, Heamophilus, Pasteurella), Pseudomonas, Rickettsiaceae, Chlamydiaceae, Syphilis, and Staphylococcus. These bacterial or fungal families can cause the numerous diseases or symptoms, including, but not limited to, bacteremia, endocarditis, eye infections (conjunctivitis, tuberculosis, uveitis), gingivitis, opportunistic infections (e.g., AIDS related

infections), paronychia, prosthesis-related infections, Reiter's Disease, respiratory tract infections (such as whooping cough or empyema), sepsis, Lyme Disease, Cat-Scratch Disease, Dysentery, Paratyphoid Fever, food poisoning, Typhoid, pneumonia, Gonorrhea, meningitis, Chlamydia, Syphilis, Diphtheria, Leprosy, Paratuberculosis, Tuberculosis, Lupus, Botulism, gangrene, tetanus, impetigo, Rheumatic Fever, Scarlet Fever, sexually-transmitted diseases, skin diseases (e.g., cellulitis, dermatocycoses), toxemia, urinary tract infections, and wound infections. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Moreover, parasitic agents causing disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but are not limited to the following families Amebiasis, Babesiosis, Coccidiosis, Cryptosporidiosis, Dientamoebiasis, Dourine, Ectoparasitic, Giardiasis, Helminthiasis, Leishmaniasis, Theileriasis, Toxoplasmosis, Trypanosomiasis, and Trichomonas. These parasites can cause a variety of diseases or symptoms, including, but not limited to, Scabies, Trombiculiasis, eye infections, intestinal disease (e.g., dysentery, giardiasis), liver disease, lung disease, opportunistic infections (e.g., AIDS related), Malaria, pregnancy complications, and toxoplasmosis. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Preferably, treatment using a polypeptide or polynucleotide of the present invention could either be by administering an effective amount of a polypeptide to the patient, or by removing cells from the patient, supplying the cells with a polynucleotide of the present invention, and returning the engineered cells to the patient (*ex vivo* therapy). Moreover, the polypeptide or polynucleotide of the present invention can be used as an antigen in a vaccine to raise an immune response against infectious disease.

Regeneration

A polynucleotide or polypeptide of the present invention can be used to differentiate, proliferate, and attract cells, leading to the regeneration of tissues (see, *Science* 276:59-87 (1997)). The regeneration of tissues could be used to repair, replace, or protect tissue damaged by congenital defects, trauma (wounds, burns, incisions, or ulcers), age, disease (e.g. osteoporosis, osteocarthritis, periodontal disease, liver failure), surgery (including cosmetic plastic surgery), fibrosis, reperfusion injury, or systemic cytokine damage.

Tissues that could be regenerated using the present invention include organs (e.g., pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac),

vascular (including vascular endothelium), nervous, hematopoietic, and skeletal (bone, cartilage, tendon, and ligament) tissue. Preferably, regeneration occurs without scarring or with minimal scarring. Regeneration also may include angiogenesis.

Moreover, a polynucleotide or polypeptide of the present invention may increase regeneration of tissues difficult to heal. For example, increased tendon/ligament regeneration would quicken recovery time after damage. A polynucleotide or polypeptide of the present invention could also be used prophylactically in an effort to avoid damage. Specific diseases that could be treated include of tendinitis, carpal tunnel syndrome, and other tendon or ligament defects. A further example of tissue regeneration of non-healing wounds includes pressure ulcers, ulcers associated with vascular insufficiency, surgical, and traumatic wounds.

Similarly, nerve and brain tissue could also be regenerated by using a polynucleotide or polypeptide of the present invention to proliferate and differentiate nerve cells. Diseases that could be treated using this method include central and peripheral nervous system diseases, neuropathies, or mechanical and traumatic disorders (e.g., spinal cord disorders, head trauma, cerebrovascular disease, and stroke). Specifically, diseases associated with peripheral nerve injuries, peripheral neuropathy (e.g., resulting from chemotherapy or other medical therapies), localized neuropathies, and central nervous system diseases (e.g., Alzheimer's disease, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome), could all be treated using the polynucleotide or polypeptide of the present invention.

Chemotaxis

A polynucleotide or polypeptide of the present invention may have chemotaxis activity. A chemotaxic molecule attracts or mobilizes cells (e.g., monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells) to a particular site in the body, such as inflammation, infection, or site of hyperproliferation. The mobilized cells can then fight off and/or heal the particular trauma or abnormality.

A polynucleotide or polypeptide of the present invention may increase chemotaxic activity of particular cells. These chemotactic molecules can then be used to treat inflammation, infection, hyperproliferative disorders, or any immune system disorder by increasing the number of cells targeted to a particular location in the body. For example, chemotaxic molecules can be used to treat wounds and other trauma to tissues by attracting immune cells to the injured location. Chemotactic molecules of the present invention can also attract fibroblasts, which can be used to treat wounds.

It is also contemplated that a polynucleotide or polypeptide of the present invention may inhibit chemotactic activity. Such molecules could also be used to treat a variety of disorders. Thus, a polynucleotide or polypeptide of the present invention could be used as an inhibitor of chemotaxis.

Binding Activity

A polypeptide of the present invention may be used to screen for molecules that bind to the polypeptide or for molecules to which the polypeptide binds. The binding of the polypeptide and the molecule may activate (i.e., an agonist), increase, inhibit (i.e., an antagonist), or decrease activity of the polypeptide or the molecule bound. Examples of such molecules include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

Preferably, the molecule is closely related to the natural ligand of the polypeptide, e.g., a fragment of the ligand, or a natural substrate, a ligand, a structural or functional mimetic (see, Coligan et al., *Current Protocols in Immunology* 1(2), Chapter 5 (1991)). Similarly, the molecule can be closely related to the natural receptor to which the polypeptide binds or, at least, a fragment of the receptor capable of being bound by the polypeptide (e.g., an active site). In either case, the molecule can be rationally designed using known techniques.

Preferably, the screening for these molecules involves producing appropriate cells which express the polypeptide, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or *E. coli*. Cells expressing the polypeptide (or cell membrane containing the expressed polypeptide) are then preferably contacted with a test compound potentially containing the molecule to observe binding, stimulation, or inhibition of activity of either the polypeptide or the molecule.

The assay may simply test binding of a candidate compound to the polypeptide, wherein binding is detected by a label, or in an assay involving competition with a labeled competitor. Further, the assay may test whether the candidate compound results in a signal generated by binding to the polypeptide.

Alternatively, the assay can be carried out using cell-free preparations, polypeptide/molecule affixed to a solid support, chemical libraries, or natural product mixtures. The assay may also simply comprise the steps of mixing a candidate compound with a solution

containing a polypeptide, measuring polypeptide/molecule activity or binding, and comparing the polypeptide/molecule activity or binding to a standard.

Preferably, an ELISA assay can measure polypeptide level or activity in a sample (e.g., biological sample) using a monoclonal or polyclonal antibody. The antibody can measure polypeptide level or activity by either binding, directly or indirectly, to the polypeptide or by competing with the polypeptide for a substrate.

All of these above assays can be used as diagnostic or prognostic markers. The molecules discovered using these assays can be used to treat disease or to bring about a particular result in a patient (e.g., blood vessel growth) by activating or inhibiting the polypeptide/molecule. Moreover, the assays can discover agents which may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues.

Therefore, the invention includes a method of identifying compounds which bind to a polypeptide of the invention comprising the steps of: (a) incubating a candidate binding compound with a polypeptide of the invention; and (b) determining if binding has occurred. Moreover, the invention includes a method of identifying agonists/antagonists comprising the steps of: (a) incubating a candidate compound with a polypeptide of the invention, (b) assaying a biological activity, and (c) determining if a biological activity of the polypeptide has been altered.

Other Activities

A polypeptide or polynucleotide of the present invention may also increase or decrease the differentiation or proliferation of embryonic stem cells from a lineage other than the above described hemopoietic cells.

A polypeptide or polynucleotide of the present invention may also be used to modulate mammalian characteristics, such as body height, weight, hair color, eye color, skin, percentage of adipose tissue, pigmentation, size, and shape (e.g., cosmetic surgery). Similarly, a polypeptide or polynucleotide of the present invention may be used to modulate mammalian metabolism affecting catabolism, anabolism, processing, utilization, and storage of energy.

A polypeptide or polynucleotide of the present invention may be used to change a mammal's mental state or physical state by influencing biorhythms, circadian rhythms, depression (including depressive disorders), tendency for violence, tolerance for pain, the response to opiates and opioids, tolerance to opiates and opioids, withdrawal from opiates and

opioids, reproductive capabilities (preferably by activin or inhibin-like activity), hormonal or endocrine levels, appetite, libido, memory, stress, or other cognitive qualities.

A polypeptide or polynucleotide of the present invention may also be used as a food additive or preservative, such as to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional components.

Other Preferred Embodiments

Where a polynucleotide of the invention is down-regulated and exacerbates a pathological condition, such as ataxia telangiectasia, the expression of the polynucleotide can be increased or the level of the intact polypeptide product can be increased in order to treat, prevent, ameliorate, or modulate the pathological condition. This can be accomplished by, for example, administering a polynucleotide or polypeptide of the invention to the mammalian subject.

A polynucleotide of the invention can be administered to a mammalian subject by a recombinant expression vector comprising the polynucleotide. A mammalian subject can be a human, baboon, chimpanzee, macaque, cow, horse, sheep, pig, horse, dog, cat, rabbit, guinea pig, rat or mouse. Preferably, the recombinant vector comprises a polynucleotide shown in SEQ ID NOs:1-23 or a polynucleotide which is at least 98% identical to a nucleic acid sequence shown in SEQ ID NOs:1-23. Also, preferably, the recombinant vector comprises a variant polynucleotide that is at least 80%, 90%, or 95% identical to a polynucleotide comprising SEQ ID NOs:1-23.

The administration of a polynucleotide or recombinant expression vector of the invention to a mammalian subject can be used to express a polynucleotide in said subject for the treatment of, for example, cancer or ataxia telangiectasia. Expression of a polynucleotide in target cells, including but not limited to neurons, tumors, and muscle cells, would effect greater production of the encoded polypeptide.

There are available to one skilled in the art multiple viral and non-viral methods suitable for introduction of a nucleic acid molecule into a target cell, as described above. In addition, a naked polynucleotide can be administered to target cells. Polynucleotides and recombinant expression vectors of the invention can be administered as a pharmaceutical composition. Such

a composition comprises an effective amount of a polynucleotide or recombinant expression vector, and a pharmaceutically acceptable formulation agent selected for suitability with the mode of administration. Suitable formulation materials preferably are non-toxic to recipients at the concentrations employed and can modify, maintain, or preserve, for example, the pH, osmolarity, viscosity, clarity, color, isotonicity, odor, sterility, stability, rate of dissolution or release, adsorption, or penetration of the composition. *See Remington's Pharmaceutical Sciences* (18^{th} Ed., A.R. Gennaro, ed., Mack Publishing Company 1990).

The pharmaceutically active compounds (i.e., a polynucleotide or a vector) can be processed in accordance with conventional methods of pharmacy to produce medicinal agents for administration to patients, including humans and other mammals. Thus, the pharmaceutical composition comprising a polynucleotide or a recombinant expression vector may be made up in a solid form (including granules, powders or suppositories) or in a liquid form (*e.g.*, solutions, suspensions, or emulsions).

The dosage regimen for treating a disease with a composition comprising a polynucleotide or expression vector is based on a variety of factors, including the type or severity of the cancer or AT, the age, weight, sex, medical condition of the patient, the route of administration, and the particular compound employed. Thus, the dosage regimen may vary widely, but can be determined routinely using standard methods. A typical dosage may range from about 0.1 mg/kg to about 100 mg/kg or more, depending on the factors mentioned above.

The frequency of dosing will depend upon the pharmacokinetic parameters of the polynucleotide or vector in the formulation being used. Typically, a clinician will administer the composition until a dosage is reached that achieves the desired effect. The composition may therefore be administered as a single dose, as two or more doses (which may or may not contain the same amount of the desired molecule) over time, or as a continuous infusion via implantation device or catheter. Further refinement of the appropriate dosage is routinely made by those of ordinary skill in the art and is within the ambit of tasks routinely performed by them. Appropriate dosages may be ascertained through use of appropriate dose-response data.

The cells of a mammalian subject may be transfected *in vivo*, *ex vivo*, or *in vitro*. Administration of a polynucleotide or a recombinant vector containing a polynucleotide to a target cell *in vivo* may be accomplished using any of a variety of techniques well known to

those skilled in the art. For example, U.S. Patent No. 5,672,344 describes an *in vivo* viral-mediated gene transfer system involving a recombinant neurotrophic HSV-1 vector. The above-described compositions of polynucleotides and recombinant vectors can be transfected *in vivo* by oral, buccal, parenteral, rectal, or topical administration as well as by inhalation spray. The term "parenteral" as used herein includes, subcutaneous, intravenous, intramuscular, intrasternal, infusion techniques or intraperitoneally.

While the nucleic acids and/or vectors of the invention can be administered as the sole active pharmaceutical agent, they can also be used in combination with one or more vectors of the invention or other agents. When administered as a combination, the therapeutic agents can be formulated as separate compositions that are given at the same time or different times, or the therapeutic agents can be given as a single composition.

Another delivery system for polynucleotides of the invention is a "non-viral" delivery system. Techniques that have been used or proposed for gene therapy include DNA-ligand complexes, adenovirus-ligand-DNA complexes, direct injection of DNA, $CaPO_4$ precipitation, gene gun techniques, electroporation, lipofection, and colloidal dispersion (Mulligan, R., (1993) *Science*, 260 (5110):926-32). Any of these methods are widely available to one skilled in the art and would be suitable for use in the present invention. Other suitable methods are available to one skilled in the art, and it is to be understood that the present invention may be accomplished using any of the available methods of transfection. Several such methodologies have been utilized by those skilled in the art with varying success (Mulligan, R., (1993) *Science*, 260 (5110):926-32).

Where a polynucleotide of the invention is up-regulated and exacerbates a pathological condition in a mammalian subject, such as ataxia telangiectasia, the expression of the polynucleotide can be blocked or reduced or the level of the intact polypeptide product can be reduced in order to treat, prevent, ameliorate, or modulate the pathological condition. This can be accomplished by, for example, the use of antisense oligonucleotides or ribozymes. Alternatively, drugs or antibodies that bind to and inactivate the polypeptide product can be used.

Antisense oligonucleotides are nucleotide sequences which are complementary to a specific DNA or RNA sequence. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form complexes and block either transcription or translation. Preferably, an antisense oligonucleotide is at least 11 nucleotides in length, but can be at least 12, 15, 20, 25, 30, 35, 40, 45, or 50 or more nucleotides long. Longer sequences also can be used. Antisense oligonucleotide molecules can be provided in a DNA construct and introduced into a cell as described above to decrease the level of gene products of the invention in the cell.

Antisense oligonucleotides can be deoxyribonucleotides, ribonucleotides, or a combination of both. Oligonucleotides can be synthesized manually or by an automated synthesizer, by covalently linking the 5' end of one nucleotide with the 3' end of another nucleotide with non-phosphodiester internucleotide linkages such alkylphosphonates, phosphorothioates, phosphorodithioates, alkylphosphonothioates, alkylphosphonates, phosphoramidates, phosphate esters, carbamates, acetamidate, carboxymethyl esters, carbonates, and phosphate triesters. *See* Brown, (1994) *Meth. Mol. Biol.*, 20:1-8; Sonveaux, (1994) *Meth. Mol. Biol.*, 26:1-72; Uhlmann et al., (1990) *Chem. Rev.*, 90:543-583.

Modifications of gene expression can be obtained by designing antisense oligonucleotides which will form duplexes to the control, 5', or regulatory regions of a gene of the invention. Oligonucleotides derived from the transcription initiation site, *e.g.*, between positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or chaperons. Therapeutic advances using triplex DNA have been described in the literature (*e.g.*, Gee et al., in Huber & Carr, MOLECULAR AND IMMUNOLOGIC APPROACHES, Futura Publishing Co., Mt. Kisco, N.Y., 1994). An antisense oligonucleotide also can be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Precise complementarity is not required for successful complex formation between an antisense oligonucleotide and the complementary sequence of a polynucleotide. Antisense oligonucleotides which comprise, for example, 2, 3, 4, or 5 or more stretches of contiguous

nucleotides which are precisely complementary to a polynucleotide, each separated by a stretch of contiguous nucleotides which are not complementary to adjacent nucleotides, can provide sufficient targeting specificity for mRNA. Preferably, each stretch of complementary contiguous nucleotides is at least 4, 5, 6, 7, or 8 or more nucleotides in length. Non-complementary intervening sequences are preferably 1, 2, 3, or 4 nucleotides in length. One skilled in the art can easily use the calculated melting point of an antisense-sense pair to determine the degree of mismatching which will be tolerated between a particular antisense oligonucleotide and a particular polynucleotide sequence.

Antisense oligonucleotides can be modified without affecting their ability to hybridize to a polynucleotide of the invention. These modifications can be internal or at one or both ends of the antisense molecule. For example, internucleoside phosphate linkages can be modified by adding cholesteryl or diamine moieties with varying numbers of carbon residues between the amino groups and terminal ribose. Modified bases and/or sugars, such as arabinose instead of ribose, or a 3', 5'-substituted oligonucleotide in which the 3' hydroxyl group or the 5' phosphate group are substituted, also can be employed in a modified antisense oligonucleotide. These modified oligonucleotides can be prepared by methods well known in the art. *See, e.g.*, Agrawal et al., (1992) *Trends Biotechnol.*, 10:152-158; Uhlmann et al., (1990) *Chem. Rev.*, 90:543-584; Uhlmann et al., (1987) *Tetrahedron. Lett.*, 215:3539-3542.

Ribozymes are RNA molecules with catalytic activity. *See, e.g.*, Cech, (1987) *Science*, 236:1532-1539; Cech, (1990) *Ann. Rev. Biochem.*, 59:543-568; Cech, (1992) *Curr. Opin. Struct. Biol.*, 2:605-609; Couture & Stinchcomb, (1996) *Trends Genet.*, 12:510-515. Ribozymes can be used to inhibit gene function by cleaving an RNA sequence, as is known in the art (*e.g.*, Haseloff et al., U.S. Patent 5,641,673). The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. Examples include engineered hammerhead motif ribozyme molecules that can specifically and efficiently catalyze endonucleolytic cleavage of specific nucleotide sequences.

The coding sequence of a polynucleotide of the invention can be used to generate ribozymes which will specifically bind to mRNA transcribed from the polynucleotide. Methods of designing and constructing ribozymes which can cleave RNA molecules in *trans* in a highly sequence specific manner have been developed and described in the art (*see* Haseloff

et al. (1988) *Nature,* 334:585-591). For example, the cleavage activity of ribozymes can be targeted to specific RNAs by engineering a discrete "hybridization" region into the ribozyme. The hybridization region contains a sequence complementary to the target RNA and thus specifically hybridizes with the target (*see, e.g.,* Gerlach et al., EP 321,201).

Specific ribozyme cleavage sites within a RNA target can be identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the target RNA containing the cleavage site can be evaluated for secondary structural features which may render the target inoperable. Suitability of candidate RNA targets also can be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays. The nucleotide sequences shown in SEQ ID NOs:1-23 and their complements provide sources of suitable hybridization region sequences. Longer complementary sequences can be used to increase the affinity of the hybridization sequence for the target. The hybridizing and cleavage regions of the ribozyme can be integrally related such that upon hybridizing to the target RNA through the complementary regions, the catalytic region of the ribozyme can cleave the target.

Ribozymes can be introduced into cells as part of a DNA construct. Mechanical methods, such as microinjection, liposome-mediated transfection, electroporation, or calcium phosphate precipitation, can be used to introduce a ribozyme-containing DNA construct into cells in which it is desired to decrease polynucleotide expression. Alternatively, if it is desired that the cells stably retain the DNA construct, the construct can be supplied on a plasmid and maintained as a separate element or integrated into the genome of the cells, as is known in the art. A ribozyme-encoding DNA construct can include transcriptional regulatory elements, such as a promoter element, an enhancer or UAS element, and a transcriptional terminator signal, for controlling transcription of ribozymes in the cells.

As taught in Haseloff et al., U.S. Patent 5,641,673, ribozymes can be engineered so that ribozyme expression will occur in response to factors which induce expression of a target gene. Ribozymes also can be engineered to provide an additional level of regulation, so that destruction of mRNA occurs only when both a ribozyme and a target gene are induced in the cells.

Pathological conditions or susceptibility to pathological conditions, such as ataxia telangiectasia, can be diagnosed using methods of the invention. Testing for expression of a polynucleotide of the invention or for the presence of the polynucleotide product can correlate with the severity of the condition and can also indicate appropriate treatment. For example, the presence or absence of a mutation in a polynucleotide of the invention can be determined and a pathological condition or a susceptibility to a pathological condition is diagnosed based on the presence or absence of the mutation. Further, an alteration in expression of a polypeptide encoded by a polynucleotide of the invention can be detected, where the presence of an alteration in expression of the polypeptide is indicative of the pathological condition or susceptibility to the pathological condition. The alteration in expression can be an increase in the amount of expression or a decrease in the amount of expression.

As an additional method of diagnosis, a first biological sample from a patient suspected of having a pathological condition, such as ataxia telangiectasia, is obtained along with a second sample from a suitable comparable control source. A biological sample can comprise saliva, blood, cerebrospinal fluid, amniotic fluid, urine, feces, or tissue, such as gastrointestinal tissue. A suitable control source can be obtained from one or more mammalian subjects that do not have the pathological condition. For example, the average concentrations and distribution of a polynucleotide or polypeptide of the invention can be determined from biological samples taken from a representative population of mammalian subjects, wherein the mammalian subjects are the same species as the subject from which the test sample was obtained. The amount of at least one polypeptide encoded by a polynucleotide of the invention is determined in the first and second sample. The amounts of the polypeptide in the first and second samples are compared. A patient is diagnosed as having a pathological condition if the amount of the polypeptide in the first sample is greater than or less than the amount of the polypeptide in the second sample.

Other preferred embodiments of the claimed invention include an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 80%, preferably at least 85%, more preferably at least 90%, most preferably at least 95% identical to a sequence of at least about 40 contiguous nucleotides in the nucleotide sequence of polynucleotides identified in SEQ ID NOs:1-23.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of polynucleotides identified in SEQ ID

NOs:1-23 in the range of positions beginning with the nucleotide at about the position of the 5' nucleotide of the clone sequence and ending with the nucleotide at about the position of the 3' nucleotide of the clone sequence.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of polynucleotides identified in SEQ ID NOs:1-23 in the range of positions beginning with the nucleotide at about the position of the 5' nucleotide of the start codon and ending with the nucleotide at about the position of the 3' nucleotide of the clone sequence as defined for polynucleotides identified in SEQ ID NOs:1-23.

Similarly preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of polynucleotides identified in SEQ ID NOs:1-23 in the range of positions beginning with the nucleotide at about the position of the 5' nucleotide of the first amino acid of the signal peptide and ending with the nucleotide at about the position of the 3' nucleotide of the clone sequence as defined for polynucleotides identified in SEQ ID NOs:1-23.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 150 contiguous nucleotides in the nucleotide sequence of polynucleotides identified in SEQ ID NOs:1-23.

Further preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 500 contiguous nucleotides in the nucleotide sequence of polynucleotides identified in SEQ ID NOs:1-23.

A further preferred embodiment is a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the nucleotide sequence of polynucleotides identified in SEQ ID NOs:1-23 beginning with the nucleotide at about the position of the 5' nucleotide of the first amino acid of the signal peptide and ending with the nucleotide at about the position of the 3' nucleotide of the clone sequence as defined for polynucleotides identified in SEQ ID NOs:1-23.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence of polynucleotides identified in SEQ ID NOs:1-23.

Also preferred is an isolated nucleic acid molecule which hybridizes under stringent hybridization conditions to a nucleic acid molecule, wherein said nucleic acid molecule which hybridizes does not hybridize under stringent hybridization conditions to a nucleic acid molecule having a nucleotide sequence consisting of only A residues or of only T residues.

A further preferred embodiment is a method for detecting in a biological sample a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 40 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of polynucleotides identified in SEQ ID NOs:1-23, which method comprises a step of comparing a nucleotide sequence of at least one nucleic acid molecule in said sample with a sequence selected from said group and determining whether the sequence of said nucleic acid molecule in said sample is at least 95% identical to said selected sequence.

Also preferred is the above method wherein said step of comparing sequences comprises determining the extent of nucleic acid hybridization between nucleic acid molecules in said sample and a nucleic acid molecule comprising said sequence selected from said group. Similarly, also preferred is the above method wherein said step of comparing sequences is performed by comparing the nucleotide sequence determined from a nucleic acid molecule in said sample with said sequence selected from said group. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

A further preferred embodiment is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting nucleic acid molecules in said sample, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 40 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of polynucleotides identified in SEQ ID NOs:1-23.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene, which method comprises a step of detecting in a biological sample obtained from said subject nucleic acid molecules, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 40 contiguous nucleotides in a sequence selected from the group consisting of a nucleotide sequence of polynucleotides identified in SEQ ID NOs:1-23.

The method for diagnosing a pathological condition can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide

sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 40 contiguous nucleotides in a sequence selected from said group.

Also preferred is a composition of matter comprising isolated nucleic acid molecules wherein the nucleotide sequences of said nucleic acid molecules comprise a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 40 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of polynucleotides identified in SEQ ID NOs:1-23. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in an amino acid sequence translated from polynucleotides identified in SEQ ID NOs:1-23.

Also preferred is a polypeptide, wherein said sequence of contiguous amino acids is included in amino acids in an amino acid sequence translated from polynucleotides identified in SEQ ID NOs:1-23, in the range of positions beginning with the residue at about the position of the first amino acid of the secreted portion and ending with the residue at about the last amino acid of the open reading frame.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in an amino acid sequence translated from polynucleotides identified in SEQ ID NOs:1-23.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in an amino acid sequence translated from polynucleotides identified in SEQ ID NOs:1-23.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to amino acids in an amino acid sequence translated from polynucleotides identified in SEQ ID NOs:1-23.

Further preferred is a method for detecting in a biological sample a polypeptide comprising an amino acid sequence which is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of amino acid sequences translated from polynucleotides identified in SEQ ID NOs:1-23, which method

comprises a step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group and determining whether the sequence of said polypeptide molecule in said sample is at least 90% identical to said sequence of at least 10 contiguous amino acids.

Also preferred is the above method wherein said step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group comprises determining the extent of specific binding of polypeptides in said sample to an antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of amino acid sequences translated from polynucleotides identified in SEQ ID NOs:1-23.

Also preferred is the above method wherein said step of comparing sequences is performed by comparing the amino acid sequence determined from a polypeptide molecule in said sample with said sequence selected from said group.

Also preferred is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting polypeptide molecules in said sample, if any, comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of amino acid sequences translated from polynucleotides identified in SEQ ID NOs:1-23.

Also preferred is the above method for identifying the species, tissue or cell type of a biological sample, which method comprises a step of detecting polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the above group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene, which method comprises a step of detecting in a biological sample obtained from said subject polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of amino acid sequences translated from polynucleotides identified in SEQ ID NOs:1-23.

In any of these methods, the step of detecting said polypeptide molecules includes using an antibody.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a nucleotide sequence encoding a polypeptide wherein said polypeptide comprises an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of amino acid sequences translated from polynucleotides identified in SEQ ID NOs:1-23.

Also preferred is an isolated nucleic acid molecule, wherein said nucleotide sequence encoding a polypeptide has been optimized for expression of said polypeptide in a prokaryotic host.

Also preferred is an isolated nucleic acid molecule, wherein said nucleotide sequence encodes a polypeptide comprising an amino acid sequence selected from the group consisting of amino acid sequences translated from polynucleotides identified in SEQ ID NOs:1-23.

Further preferred is a method of making a recombinant vector comprising inserting any of the above isolated nucleic acid molecule into a vector. Also preferred is the recombinant vector produced by this method. Also preferred is a method of making a recombinant host cell comprising introducing the vector into a host cell, as well as the recombinant host cell produced by this method.

Also preferred is a method of making an isolated polypeptide comprising culturing this recombinant host cell under conditions such that said polypeptide is expressed and recovering said polypeptide. Also preferred is this method of making an isolated polypeptide, wherein said recombinant host cell is a eukaryotic cell and said polypeptide is a secreted portion of a human secreted protein comprising an amino acid sequence selected from the group consisting of amino acid sequences translated from polynucleotides identified in SEQ ID NOs:1-23. The isolated polypeptide produced by this method is also preferred.

Also preferred is a method of treatment of an individual in need of an increased level of a secreted protein activity, which method comprises administering to such an individual a pharmaceutical composition comprising an amount of an isolated polypeptide, polynucleotide, or antibody of the claimed invention effective to increase the level of said protein activity in said individual.

The present invention also includes a diagnostic system, preferably in kit form, for assaying for the presence of the polypeptide of the present invention in a body sample, such as brain tissue, cell suspensions or tissue sections, or body fluid samples such as CSF, blood, plasma or serum, where it is desirable to detect the presence, and preferably the amount, of the polypeptide of this invention in the sample according to the diagnostic methods described herein.

In a related embodiment, a nucleic acid molecule can be used as a probe (i.e., an oligonucleotide) to detect the presence of a polynucleotide of the present invention, a gene corresponding to a polynucleotide of the present invention, or a mRNA in a cell that is diagnostic for the presence or expression of a polypeptide of the present invention in the cell. The nucleic acid molecule probes can be of a variety of lengths from at least about 10 contiguous bases, suitably about 10 to about 5000 nucleotides long, although they will typically be about 20 to 500 nucleotides in length. Hybridization methods are extremely well known in the art and will not be described further here.

In a related embodiment, detection of genes corresponding to the polynucleotides of the present invention can be conducted by primer extension reactions such as the polymerase chain reaction (PCR). To that end, PCR primers are utilized in pairs, as is well known, based on the nucleotide sequence of the gene to be detected. Preferably, the nucleotide sequence is a portion of the nucleotide sequence of a polynucleotide of the present invention. Particularly preferred PCR primers can be derived from any portion of a DNA sequence encoding a polypeptide of the present invention, but are preferentially from regions which are not conserved in other cellular proteins.

Preferred PCR primer pairs useful for detecting the genes corresponding to the polynucleotides of the present invention and expression of these genes are described in the Examples, including the corresponding Tables. Nucleotide primers from the corresponding region of the polypeptides of the present invention described herein are readily prepared and used as PCR primers for detection of the presence or expression of the corresponding gene in any of a variety of tissues.

The diagnostic system includes, in an amount sufficient to perform at least one assay, a subject polypeptide of the present invention, a subject antibody or monoclonal antibody, and/or

a subject nucleic acid molecule probe of the present invention, as a separately packaged reagent.

In another embodiment, a diagnostic system, preferably in kit form, is contemplated for assaying for the presence of the polypeptide of the present invention or an antibody immunoreactive with the polypeptide of the present invention in a body fluid sample. Such diagnostic kit would be useful for monitoring the fate of therapeutically administered polypeptide of the present invention or an antibody immunoreactive with the polypeptide of the present invention. The system includes, in an amount sufficient for at least one assay, a polypeptide of the present invention and/or a subject antibody as a separately packaged immunochemical reagent.

Instructions for use of the packaged reagent(s) are also typically included.

As used herein, the term "package" refers to a solid matrix or material such as glass, plastic (e.g., polyethylene, polypropylene or polycarbonate), paper, foil and the like capable of holding within fixed limits a polypeptide, polyclonal antibody or monoclonal antibody of the present invention. Thus, for example, a package can be a glass vial used to contain milligram quantities of a contemplated polypeptide or antibody or it can be a microtiter plate well to which microgram quantities of a contemplated polypeptide or antibody have been operatively affixed, (i.e., linked) so as to be capable of being immunologically bound by an antibody or antigen, respectively.

"Instructions for use" typically include a tangible expression describing the reagent concentration or at least one assay method parameter such as the relative amounts of reagent and sample to be admixed, maintenance time periods for reagent/ sample admixtures, temperature, buffer conditions and the like.

A diagnostic system of the present invention preferably also includes a label or indicating means capable of signaling the formation of an immunocomplex containing a polypeptide or antibody molecule of the present invention.

The word "complex" as used herein refers to the product of a specific binding reaction such as an antibody-antigen or receptor-ligand reaction. Exemplary complexes are immunoreaction products.

As used herein, the terms "label" and "indicating means" in their various grammatical forms refer to single atoms and molecules that are either directly or indirectly involved in the production of a detectable signal to indicate the presence of a complex. Any label or indicating means can be linked to or incorporated in an expressed protein, polypeptide, or antibody molecule that is part of an antibody or monoclonal antibody composition of the present invention, or used separately, and those atoms or molecules can be used alone or in conjunction with additional reagents. Such labels are themselves well-known in clinical diagnostic chemistry and constitute a part of this invention only insofar as they are utilized with otherwise novel proteins methods and/or systems.

The labeling means can be a fluorescent labeling agent that chemically binds to antibodies or antigens without denaturing them to form a fluorochrome (dye) that is a useful immunofluorescent tracer. Suitable fluorescent labeling agents are fluorochromes such as fluorescein isocyanate (FIC), fluorescein isothiocyante (FITC), 5-dimethylamine-1-naphthalenesulfonyl chloride (DANSC), tetramethylrhodamine isothiocyanate (TRITC), lissamine, rhodamine 8200 sulphonyl chloride (RB 200 SC) and the like. A description of immunofluorescence analysis techniques is found in DeLuca, "Immunofluorescence Analysis", in *Antibody As a Tool*, Marchalonis, et al., Eds., John Wiley & Sons, Ltd., pp. 189-231 (1982), which is incorporated herein by reference. Other suitable labeling agents are known to those skilled in the art.

In preferred embodiments, the indicating group is an enzyme, such as horseradish peroxidase (HRP), glucose oxidase, or the like. In such cases where the principal indicating group is an enzyme such as HRP or glucose oxidase, additional reagents are required to visualize the fact that a receptor-ligand complex (immunoreactant) has formed. Such additional reagents for HRP include hydrogen peroxide and an oxidation dye precursor such as diaminobenzidine. An additional reagent useful with glucose oxidase is 2,2'-amino-di-(3-ethyl-benzthiazoline-G-sulfonic acid) (ABTS).

Radioactive elements are also useful labeling agents and are used illustratively herein. An exemplary radiolabeling agent is a radioactive element that produces gamma ray emissions. Elements which themselves emit gamma rays, such as ^{124}I, ^{125}I, ^{128}I, ^{132}I and ^{51}Cr represent one class of gamma ray emission-producing radioactive element indicating groups. Particularly preferred is ^{125}I. Another group of useful labeling means are those elements such as ^{11}C, ^{18}F, ^{15}O and ^{13}N which themselves emit positrons. The positrons so emitted produce gamma rays upon encounters with electrons present in the animal's body. Also useful is a beta emitter, such ^{111} indium or ^{3}H.

The linking of labels, and the labeling of polypeptides and proteins is well known in the art. For instance, antibody molecules produced by a hybridoma can be labeled by metabolic incorporation of radioisotope-containing amino acids provided as a component in the culture medium. See, e.g., Galfre et al., *Meth. Enzymol.*, 73:3-46 (1981). The techniques of protein conjugation or coupling through activated functional groups are particularly applicable. See, e.g., Aurameas, et al., *Scand. J. Immunol.*, Vol. 8 Suppl. 7:7-23 (1978), Rodwell et al., *Biotech.*, 3:889-894 (1984), and U.S. Pat. No. 4,493,795.

The diagnostic systems can also include, preferably as a separate package, a specific binding agent. A "specific binding agent" is a molecular entity capable of selectively binding a reagent species of the present invention or a complex containing such a species, but is not itself a polypeptide or antibody molecule composition of the present invention. Exemplary specific binding agents are second antibody molecules, complement proteins or fragments thereof, *S. aureus* protein A, and the like. Preferably the specific binding agent binds the reagent species when that species is present as part of a complex.

In preferred embodiments, the specific binding agent is labeled. However, when the diagnostic system includes a specific binding agent that is not labeled, the agent is typically used as an amplifying means or reagent. In these embodiments, the labeled specific binding agent is capable of specifically binding the amplifying means when the amplifying means is bound to a reagent species-containing complex.

The diagnostic kits of the present invention can be used in an "ELISA" format to detect the quantity of the polypeptide of the present invention in a sample. "ELISA" refers to an enzyme-linked immunosorbent assay that employs an antibody or antigen bound to a solid

phase and an enzyme-antigen or enzyme-antibody conjugate to detect and quantify the amount of an antigen present in a sample. A description of the ELISA technique is found in Sites et al., *Basic and Clinical Immunology*, 4th Ed. Chap. 22, Lange Medical Publications, Los Altos, CA (1982) and in U.S. Patent No. 3,654,090; U.S. Patent No. 3,850,752; and U.S. Patent No. 4,016,043, which are all incorporated herein by reference.

Thus, in some embodiments, a polypeptide of the present invention, an antibody or a monoclonal antibody of the present invention can be affixed to a solid matrix to form a solid support that comprises a package in the subject diagnostic systems.

A reagent is typically affixed to a solid matrix by adsorption from an aqueous medium, although other modes of affixation applicable to proteins and polypeptides can be used that are well known to those skilled in the art. Exemplary adsorption methods are described herein.

Useful solid matrices are also well known in the art. Such materials are water insoluble and include the cross-linked dextran available under the trademark SEPHADEX from Pharmacia Fine Chemicals (Piscataway, NJ), agarose, polystyrene beads of about 1 micron (μm) to about 5 millimeters (mm) in diameter available from several suppliers, (e.g., Abbott Laboratories, Chicago, IL), polyvinyl chloride, polystyrene, cross-linked polyacrylamide, nitrocellulose- or nylon-based webs (sheets, strips or paddles) or tubes, plates or the wells of a microtiter plate, such as those made from polystyrene or polyvinylchloride.

The reagent species, labeled specific binding agent, or amplifying reagent of any diagnostic system described herein can be provided in solution, as a liquid dispersion or as a substantially dry power, e.g., in lyophilized form. Where the indicating means is an enzyme, the enzyme's substrate can also be provided in a separate package of a system. A solid support such as the before-described microtiter plate and one or more buffers can also be included as separately packaged elements in this diagnostic assay system.

The packaging materials discussed herein in relation to diagnostic systems are those customarily utilized in diagnostic systems.

Having generally described the invention, the same will be more readily understood by reference to the following examples, which are provided by way of illustration and are not

intended as limiting.

EXAMPLE 1

Identification and Characterization of Regulated Polynucleotides

Spectral Karyotyping (SKY) and Fluorescent In Situ Hybridization (FISH) of *Atm* deficient T-cell lymphoblastic lymphomas

Ten different *Atm* deficient tumor cell lines were established. The *Atm* deficient tumor cell lines have been analyzed with a variety of methods and shown to maintain their clonality in culture even with continuous passage. These cell lines were used as a source of mRNA and were characterized further. The genetic alterations in these cell lines were characterized using spectral karyotyping (SKY) and fluorescent *in situ* hybridization (FISH). Consistent abnormalities of chromosome 14 and chromosome 12, similar to those found in human leukemias and lymphomas, were observed in all tumors studied. Examples of such abnormalities are shown in Table 1. The observed abnormalities included insertions (ins6:14) as well as translocations (t12;10, t14;15 and t14;3).

Abnormalities at the TCR locus as found in human hematopoietic cancer were observed. Genomic clones in bacterial artificial chromosomes (BACs) containing the candidate gene of interest were selected. These BACs were fluorescently labeled and used as probes on metaphase preparations using FISH. The results of these experiments demonstrated that both alleles of the TCRα locus are abnormally rearranged in all tumor cell lines studied. These results further demonstrated that the TCR locus is rearranged in both human and mouse lymphomas. In addition, two of ten tumors analyzed showed chromosomal abnormalities of genes known to be involved in human tumors.

ATM function is therefore critical for maintaining appropriate recombination pathways and in the absence of ATM, both humans and mice develop aggressive tumors. A similar pathway of tumorigenesis is also likely to occur in the absence of ATM in both human and mouse. Further studies have involved the transformation of the normal T cell to a malignant cell as described below.

Cytoplasmic RNA was prepared from cells from four tumor cell lines: AT-4, AT-7, AT-12 and AT-13 that had the chromosomal abnormalities summarized in Table 1.

Table 1

T-Cell Lymphomas From *Atm -/-* Mice Show Consistent Abnormalities of Chromosomes 12 & 14

Tumor Cell Line	No. of Chromosomal Abnormalities	Chromosomal Abnormalities	Gains/ Losses
AT-1	4	T(12;14); Del (12); T(14;15); Dup(15); T(14;3); Dup(3); Ins(6;14)	2/1
AT-2	3	T(12;15); T(15;12); T(4;3); Del(4); Dup(3)	1/1
AT-3	4	T(12;8); Del (12); T(13;10); Dup(10); Ins(2;6); Dup(6); T(8;13)	2/1
AT-4	4	T(12;14); Del (12); T(14;15); Dup(15); T(14;X); Del(14); T(X;11); Ins(6;14)	2/2
AT-5	3	T(12;14); Del (12); T(14;15); Dup(15); T(14;X); Del(14); Dup(X)	2/2
AT-7	4	T(12;14); Del (12); T(14;16); Ins(14;1); Del(14); Del(14) +15	1/3
AT-10	6	T(12;10); Del (12); Dup(14); T(X;15); Dup(15); Del (X); T(17;1); Dup(1); Del(16) -11	3/4
AT-11	5	T(12;6); Del (12); Dup(6); Dic (14;14); Dup(14); Dic (15;15); Dup(15); Rb(16;16); Dup(16) +15	5/1
AT-12	6	T(12:9); Del (12); T(9;15); Dup(15); Del(14) +10 +13 +15	4/2
AT-13	3	T(12;14); Del (12); T(14;15); Ins(14;15); Dup(15); Del(14)	1/2
(T-translocation, Dup-duplication, Del-deletion, Ins-insertion)			

Isolated RNA was analyzed using a method of simultaneous sequence-specific identification of mRNAs known as TOGA (TOtal Gene expression Analysis) described in Sutcliffe, J.G. et al., *Proc. Natl. Acad. Sci. USA*, 97(5):1976-1981 (2000); International published application PCT/US99/23655, U.S. Patent No. 5,459,037, U.S. Patent No. 5,807,680, U.S. Patent No. 6,030,784, and U.S. Patent No. 6,096,503, hereby incorporated herein by reference. Preferably, prior to the application of the TOGA technique, the isolated RNA is enriched to form a starting polyA-containing mRNA population by methods known in the art. In such a preferred embodiment, the TOGA method further comprises an additional Polymerase Chain Reaction ("PCR") step performed using four 5' PCR primers in four separate reactions and cDNA templates prepared from a population of antisense cRNAs. A final PCR step that used 256 5' PCR primers in separate reactions produced PCR products that were cDNA fragments that corresponded to a 3'-region of the starting mRNA population. The produced PCR products were then identified by a) the initial 5' sequence comprising the sequence remainder of the recognition site of the restriction endonuclease used to cut and isolate the 3' region plus the sequence of the preferably four parsing bases immediately 3' to the remainder of the recognition site, preferably the sequence of the entire fragment, and b) the length of the fragment. These two parameters, sequence and fragment length, were used to compare the

obtained PCR products to a database of known polynucleotide sequences. Since the length of the obtained PCR products includes known vector sequences at the 5' and 3' ends of the insert, the sequence of the insert provided in the sequence listing is shorter than the fragment length that forms part of the digital address.

The method yields Digital Sequence Tags (DSTs), that is, polynucleotides that are expressed sequence tags (ESTs) of the 3' end of mRNAs. DSTs that showed changes in relative levels when comparing 1 of 4 samples, or showed paired differences in 2 of 4 samples or were highly expressed and located on mouse chromosomes involved in the translocations were selected for further study. The intensities of the laser-induced fluorescence of the labeled PCR products were compared across samples isolated from the four tumor cell lines AT-4, AT-7, AT-12 and AT-13. The results are presented in Tables 2 and 3.

In general, double-stranded cDNA is generated from poly(A)-enriched cytoplasmic RNA extracted from the tissue samples of interest using an equimolar mixture of all 48 5'-biotinylated anchor primers of a set to initiate reverse transcription. One such suitable set is G-A-A-T-T-C-A-A-C-T-G-G-A-A-G-C-G-G-C-C-C-G-C-A-G-G-A-A-T-T-T-T-T-T-T-T-T-T-T-T-T-T-T-T-T-T-T-V-N-N (SEQ ID NO:24), where V is A, C or G and N is A, C, G or T. One member of this mixture of 48 anchor primers initiates synthesis at a fixed position at the 3' end of all copies of each mRNA species in the sample, thereby defining a 3' endpoint for each species, resulting in biotinylated double stranded cDNA.

Each biotinylated double stranded cDNA sample was cleaved with the restriction endonuclease MspI, which recognizes the sequence CCGG. The 3' fragments of cDNA were then isolated by capture of the biotinylated cDNA fragments on a streptavidin-coated substrate. Suitable streptavidin-coated substrates include microtitre plates, PCR tubes, polystyrene beads, paramagnetic polymer beads and paramagnetic porous glass particles. A preferred streptavidin-coated substrate is a suspension of paramagnetic polymer beads (Dynal, Inc., Lake Success, NY).

After washing the streptavidin-coated substrate and captured biotinylated cDNA fragments, the cDNA fragment product was released by digestion with NotI, which cleaves at an 8-nucleotide sequence within the anchor primers but rarely within the mRNA-derived portion of the cDNAs. The 3' MspI-NotI fragments, which are of uniform length for each

mRNA species, were directionally ligated into ClaI-, NotI-cleaved plasmid pBC SK$^+$ (Stratagene, La Jolla, CA) in an antisense orientation with respect to the vector's T3 promoter, and the product used to transform *Escherichia coli* SURE cells (Stratagene). The ligation regenerates the NotI site, but not the MspI site leaving CGG as the first three bases of the 5' end of all PCR products obtained. Each library contained in excess of 5 x 10^5 recombinants to ensure a high likelihood that the 3' ends of all mRNAs with concentrations of 0.001% or greater were multiply represented. Plasmid preps (Qiagen) were made from the cDNA library of each sample under study.

An aliquot of each library was digested with MspI, which effects linearization by cleavage at several sites within the parent vector while leaving the 3' cDNA inserts and their flanking sequences, including the T3 promoter, intact. The product was incubated with T3 RNA polymerase (MEGAscript kit, Ambion) to generate antisense cRNA transcripts of the cloned inserts containing known vector sequences abutting the MspI and NotI sites from the original cDNAs.

At this stage, each of the cRNA preparations was processed in a three-step fashion. In step one, 250 ng of cRNA was converted to first-strand cDNA using the 5' RT primer (A-G-G-T-C-G-A-C-G-G-T-A-T-C-G-G (SEQ ID NO:25). In step two, 400 pg of cDNA product was used as PCR template in four separate reactions with each of the four 5' PCR primers of the form G-G-T-C-G-A-C-G-G-T-A-T-C-G-G-N (SEQ ID NO:26), each paired with a "universal" 3' PCR primer G-A-G-C-T-C-C-A-C-C-G-C-G-G-T (SEQ ID NO:27) to yield four sets of PCR reaction products ("N1 reaction products").

In step three, the product of each subpool was further divided into 64 subsubpools (2ng in 20μl) for the second PCR reaction. This PCR reaction comprised adding 100 ng of the fluoresceinated "universal" 3' PCR primer (SEQ ID NO: 27) conjugated to 6-FAM and 100 ng of the appropriate 5' PCR primer of the form C-G-A-C-G-G-T-A-T-C-G-G-N-N-N-N (SEQ ID NO: 28), and using a program that included an annealing step at a temperature X slightly above the T_m of each 5' PCR primer to minimize artifactual mispriming and promote high fidelity copying. Each polymerase chain reaction step was performed in the presence of TaqStart antibody (Clonetech).

The products ("N4 reaction products") from the final polymerase chain reaction step for each of the tissue samples were resolved on a series of denaturing DNA sequencing gels using the automated ABI Prizm 377 sequencer. Data were collected using the GeneScan software package (ABI) and normalized for amplitude and migration. Complete execution of this series of reactions generated 64 product subpools for each of the four pools established by the 5' PCR primers of the first PCR reaction, for a total of 256 product subpools for the entire 5' PCR primer set of the second PCR reaction.

The mRNA samples from each of the tumor cell lines as described above were analyzed. Table 2 is a summary of the expression levels of 1136 mRNAs determined from cDNA. These cDNA molecules are identified by their digital address, that is, a partial 5' terminus nucleotide sequence coupled with the length of the molecule, as well as the relative amount of the molecule produced at different time intervals after treatment. The 5' terminus partial nucleotide sequence is determined by the recognition site for MspI and the nucleotide sequence of the parsing bases of the 5' PCR primer used in the final PCR step. The digital length of the fragment was determined by interpolation on a standard curve, and as such, may vary plus or minus 1 or 2 base pairs from the actual length as determined by sequencing.

For example, the entry in Table 2 that describes a DNA molecule identified by the digital address MspI GAGC 197, is further characterized as having a 5' terminus partial nucleotide sequence of CGGGAGC and a digital address length of 197 b.p. The DNA molecule identified as MspI GAGC 197 was further characterized as being expressed at the lowest level in cell line AT-12, at higher levels in cell lines AT-7 and AT-4 and at the highest level in the cell line AT-13.

The data for MspI GAGC 197 shown in Table 2 and Figure 1 were generated with a 5'-PCR primer (C-G-A-C-G-G-T-A-T-C-G-G-G-A-G-C, SEQ ID NO:29) paired with the "universal" 3' primer (SEQ ID NO: 27) labeled with 6-carboxyfluorescein (6FAM, ABI) at the 5' terminus. PCR reaction products were resolved by gel electrophoresis on 4.5% acrylamide gels and fluorescence data acquired on ABI377 automated sequencers. Data were analyzed using GeneScan software (Perkin-Elmer).

Similarly, the other DNA molecules identified in Table 2 by their MspI digital addresses are further characterized by the pattern of the level of gene expression in the different cell lines.

Further examples of TOGA analysis are shown in Figures 3, 5, 7, and 9. Some PCR products, which were also differentially represented, appeared to migrate in positions that suggest that the products were novel based on comparison to data extracted from GenBank. Sequence alignments are shown for selected DSTs in Table 3. Table 3 summarizes the alignment information from a BLAST report generated upon comparing the cloned sequence to those in the GenBank/EST databases. In these cases, the PCR product is isolated, cloned into a TOPO vector (Invitrogen) and sequenced on both strands. In order to verify that the clones isolated are from the same peak, PCR primers are designed based on the determined sequence and PCR is performed using the cDNA produced in the first PCR reaction as substrate. Oligonucleotides were synthesized corresponding to the 5' PCR primer in the second PCR step extended at the 3' end with an additional 14 nucleotides from the cloned sequence. The primer sequences are shown in Table 4.

TOGA analysis figures 1, 3, 5, and 7 are based on data from a single library. These experiments were replicated with a second library from the same source. Any differences between the length of the DST in the digital addresses in the single library versus the duplicate libraries are no more than two base pairs. The digital addresses reported in Table 2, the information in Tables 3 and 4, and Figures 9 and 10 are the product of an analysis that uses data obtained from both libraries.

Figure 3 is a graphical representation of the results of TOGA analysis using a 5' PCR primer with parsing bases GCTG, showing PCR products produced from mRNA extracted from (top to bottom panels) cell lines AT-4 (Fig. 3A), AT-7 (Fig. 3B), AT-12 (Fig. 3C), and AT-13 (Fig. 3D); the arrow indicates the vertical cursor connecting the position of peaks of length 345 base pairs.

Another example is shown in Figure 5, which is a graphical representation of the results of TOGA analysis using a 5' PCR primer with parsing bases ACAT, showing PCR products produced from mRNA extracted from (top to bottom panels) cell lines AT-4 (Fig. 5A), AT-7 (Fig. 5B), AT-12 (Fig. 5C), and AT-13 (Fig. 5D); the arrow indicates the vertical cursor connecting the position of peaks of length 367 base pairs.

Figure 7 shows the results of another example of TOGA analysis, in this case using a 5' PCR primer with parsing bases ACGG, showing PCR products produced from mRNA extracted

from (top to bottom panels) cell lines AT-4 (Fig. 7A), AT-7 (Fig. 7B), AT-12 (Fig. 7C), and AT-13 (Fig. 7D); the arrow indicates the vertical cursor connecting the position of peaks of length 458 base pairs.

In one case (BAR1_11; SEQ ID NO:3), the DST sequence listed was cloned into a TOPO vector as described previously. Figure10 illustrates the verification process, in which the cDNA produced in the first PCR reaction was used as template, and the PCR product obtained using an extended PCR primer designed based on the determined sequence was compared to products obtained using the original PCR primers. Oligonucleotides were synthesized corresponding to the 5' PCR primer in the second PCR step for each candidate extended at the 3' end with an additional 14 nucleotides from the cloned sequence. This clone specific 5' PCR primer (SEQ ID NO:32) was paired with a fluorescent labeled universal 3' PCR primer (SEQ ID NO:27) in PCR reactions using the cDNA produced in the first PCR reaction or substrate. Primers designed for such studies of cloned DSTs are shown in Table 4.

Figure 9 is a graphical representation of the results of TOGA analysis using a 5' PCR primer with parsing bases CCGT showing PCR products produced from mRNA extracted from cell lines AT-4 (Fig. 9A), AT-7 (Fig. 9B), AT-12 (Fig. 9C), and AT-13 (Fig. 9D); the arrow indicates the vertical cursor connecting the position of peaks of length 151 base pairs. The length of the PCR fragment amplified with an extended primer based on the cloned sequence was compared to the length of the original PCR product as shown in Figure 10. The upper panel (A) shows the extended primer reaction product created with a 5' PCR primer (SEQ ID NO:32) and the 3' PCR primer (SEQ ID NO:27). Figure 10B shows the PCR products produced using the original PCR primers. In the bottom panel (Figure 10C), the traces from the top and middle panels are overlaid, demonstrating that the PCR product produced using an extended primer based on the cloned sequence is the same length as the original PCR product.

The map locations of several DSTs were determined from database information as follows: BAR1_20 (SEQ ID NO:21) maps to chromosome 13, BAR1_24 (SEQ ID NO:5) maps to chromosome 7; 62 cM, BAR1_27 (SEQ ID NO:1) maps to chromosome 14; 20.5 cM, and BAR1_28 (SEQ ID NO:4) maps to chromosome 15; 43.3 cM.

The expression pattern of several DSTs has been assessed by Northern Blot analysis using RNA extracted from the above cell lines as well as other cell lines. RNA and multiple

northern blots have been prepared from various cell lines, including approximately 20 different tumor cell lines with mutations in *Atm*, *p53* or combinations of *Atm* and *p53*. These cell lines have been characterized using flow cytometry and the majority are T-cell lymphomas. These tumors are being characterized or have been characterized using SKY and FISH.

Several candidate genes have been chosen for evaluation based on their known localization to chromosomes that are abnormal in the specific *Atm* and *Atm*/p53 deficient tumors. Northern blots have been used to verify the gene expression patterns in the four lines subjected to TOGA. Candidate EST sequences were determined. EST sequences were used in BLAST searches to identify matching sequences in the GenBank database. These GenBank sequences were then used to generate probes as indicated. Selected examples are shown in Figures 2, 4, 6, and 8.

We have chosen to study these cell lines because of the demonstrated correlation between chromosomal aberrations in our mouse model and those of the human disease. We chose several additional cell lines for northern analysis that include tumor cell lines derived from mice with mutations in the tumor suppressor gene p53 as well as combinations of mutations of AT and p53. This will help us determine which genes are abnormally regulated due to loss of AT (cell lines labeled AT-#), p53 (p53-1)or combinations of both genes (APT-3 is AT-/-p53+/- and 101-7, 292-3 are double mutants). We will pursue genes that are abnormally regulated due to loss of AT as this is the clinical phenotype we are most interested in and for which it appears the mouse model is an ideal for determining how abnormal regulation of genes give rise to human lymphoma and leukemia.

Genes that are abnormally regulated among AT cell lines may be uniquely associated with AT tumorigenesis (as compared to other mechanisms of tumorigenesis) due to genetic abnormalities caused by AT deficiency. These may include phenomena such as translocations of lymphoid specific promoters or other regulatory sequences in or near genes promoting cell growth, or inactivation or deletion of genes that may regulate or suppress lymphoid tumor growth. Thus, the examples shown here were also chosen in part on the basis of their potential location in or near chromosomal aberrations among the AT cell lines studied. The genes identified in these examples may relate to specific biological aspects found both in the mouse model and in human AT tumorigenesis (e.g., the development of particularly aggressive forms of lymphomas).

Poly A enriched mRNA was extracted from four tumor cell lines, AT-4, AT-7, AT-12, and AT-13, as described above. The mRNA samples were electrophoresed through an agarose gel, blotted, and probed using well-known methods. Briefly, 20 µg of total RNA or 2 µg of poly A+ mRNA was electrophoresed through a 1.2% agarose gel containing formamide along with the appropriate molecular weight standards. The gel was blotted overnight using nylon membrane to transfer the RNA. The membrane was prehybridized for one hour at 42°C in hybridization buffer (5X SSPE, 5X Denhardt's solution, 50% formamide, 0.2% SDS, 100 µg/ml salmon sperm DNA, and water). The probe DNA (50ng) was labeled with ^{32}[P]-dCTP and ^{32}[P]-dATP using asymmetric PCR labeling. The membrane was probed with radiolabeled DNA (2-5 x 10^6 cpm/ ml) overnight at 42°C in hybridization buffer. In addition, the northern blots were probed with radiolabeled cyclophylin DNA to normalize the amount of mRNA in each sample. Band intensities of the probed mRNA samples were quantitated using a Phosphoimager SI and normalized to the hybridization signal of cyclophilin.

Figure 2 is a graphical representation of the results of Northern Blot analysis using radiolabeled probe derived from a candidate gene (GenBank Accession Number U88908) mapping to mouse Chromosome 9, FISH mapped to band A2 (cytogenic) (see Liston, P., (1997) *Genomics,* 46(3):495-503) corresponding to GAGC 197, showing, in the image in the upper panel, hybridization with a band of mRNA at 4.2 kb that is expressed at various levels in cell lines AT-4, AT-7, AT-10, AT-11, AT-13, APT-3, P53-1, 101-7, 292-3 and thymus, but not detectable in cell line AT12; the image in the lower panel shows the methylene blue stained gel for quantification of mRNA loading.

Figure 4 is a graphical representation of the results of Northern Blot analysis using radiolabeled probe derived from a candidate gene (GenBank Accession Number Z31664) human homolog (ACVRL1- GenBank accession Z22533) maps to 12q11-q14, this region is syntenic to mouse chromosome 15 between 56.8 cM-64.0 cM, corresponding to GCTG 345, showing, in the image in the upper panel, hybridization with bands of mRNA at 4.0 kb and 3.7 kb that are expressed at various levels in cell lines AT-4, AT-11, AT-12, AT-13, 292-3, and poly A^+ (pA) thymus mRNA, but not detectable in cell lines AT-7, AT-10, APT-3, P53-1, 101-7; the image in the lower panel shows the methylene blue stained gel for quantification of mRNA loading.

Figure 6 is a graphical representation of the results of Northern Blot analysis using radiolabeled probe derived from a candidate gene (GenBank Accession Number U38252) mapping to mouse Chromosome 5, 65.0 cM corresponding to ACAT 367, showing, in the image in the upper panel, hybridization with a band of mRNA at 4.1 kb that is expressed at various levels in cell lines AT-7, AT-10, AT-12, AT-13 and APT-3, but not detectable in cell lines AT-4, AT-11, P53-1, 101-7, 292-3 and polyA (pA) thymus mRNA; the image in the lower panel shows the methylene blue stained gel for quantification of mRNA loading.

Figure 8 is a graphical representation of the results of Northern Blot analysis using radiolabeled probe derived from a candidate gene (GenBank Accession Number X12822) mapping to mouse Chromosome 14, 20.5 cM corresponding to ACGG 458, showing, in the image in the upper panel, hybridization with bands of mRNA at 1.7 kb and 1.1 kb that are expressed at various levels in cell lines AT-7 and AT-10, but not detectable in cell lines AT-4, AT-12, AT-11, AT-13; APT-3, P53-1, 101-7, and 292-3, and polyA (pA) thymus (only 1.1 kb) mRNA; the image in the lower panel shows the methylene blue stained gel for quantification of mRNA loading.

EXAMPLE 2

Verification of Candidate Gene Targets Using FISH

Once candidate genes are identified that show altered expression in the tumor cell lines, FISH is used to map the chromosomal location of the gene. Candidate genes that fall within the region that is rearranged in the particular cell line are identified as a gene whose change in expression is essential for tumorigenesis.

TABLE 2							
Seq ID	Clone ID	Digital Address (MspI)		AT4	AT7	AT12	AT13
		AAAA	196	45	74	24	57
		AAAA	261	96	108	81	184
		AAAA	473	104	39	154	531
		AAAC	232	112	156	194	79
		AAAC	348	76	195	46	14
		AAAG	99	54	61	948	227
		AAAG	143	32	85	22	32
		AAAG	169	92	418	1381	421
		AAAG	183	28	56	79	58
		AAAG	215	23	60	48	64
		AAAG	225	404	167	148	123
		AAAG	231	294	132	134	110
		AAAG	246	140	56	48	44
		AAAG	251	100	48	36	34
		AAAG	355	98	348	312	242
		AACA	105	47	43	69	167
		AACA	127	175	99	356	212
		AACA	182	122	19	383	60
		AACA	197	52	146	55	155
		AACA	201	51	78	22	72
		AACC	235	67	37	32	18
		AACC	323	396	328	550	108
		AACC	390	55	18	14	32
		AACC	401	56	104	18	39
6	BAR1_16	AACG	119	2745	448	4106	3652
		AACG	313	32	54	150	42
		AACG	426	120	587	306	708
		AACG	430	69	565	23	408
		AACG	445	64	22	44	12
		AACT	127	134	140	66	170
		AACT	143	96	110	321	153
		AACT	159	1367	2187	1732	270
		AACT	262	384	344	136	176
		AACT	423	344	302	138	142
		AACT	474	111	28	118	108
		AAGA	108	213	395	147	136
		AAGA	171	94	178	50	257
		AAGA	234	118	44	131	129
		AAGA	325	38	16	50	60
		AAGA	333	168	226	90	238
		AAGC	165	178	615	334	624
		AAGC	300	69	47	24	36
		AAGG	140	425	590	198	444
		AAGG	251	410	582	137	432

		AAGG	375	479	158	528	484
		AAGG	438	40	52	12	33
		AAGG	480	55	24	30	22
		AAGT	403	34	36	50	14
		AAGT	441	588	776	780	478
		AATA	104	339	1532	105	626
		AATA	181	76	34	119	125
		AATA	450	54	30	50	25
		AATA	473	175	580	280	742
		AATA	491	59	29	43	26
		AATA	495	56	40	41	20
		AATC	216	521	1008	174	708
		AATC	473	103	38	50	223
		AATG	406	816	1166	294	1543
		ACAA	244	196	38	79	199
		ACAA	267	172	222	86	49
		ACAA	275	223	242	343	73
		ACAA	380	186	948	494	565
		ACAA	383	142	648	430	440
		ACAA	429	63	27	57	26
		ACAA	440	146	293	63	152
		ACAA	484	37	72	59	20
		ACAA	497	60	34	63	24
		ACAC	82	307	202	140	168
		ACAC	97	198	86	53	122
		ACAC	107	227	198	82	249
		ACAC	142	346	216	545	300
		ACAC	200	24	74	35	43
		ACAC	228	256	21	38	36
		ACAC	265	87	82	352	170
		ACAC	268	166	82	389	301
		ACAG	93	120	382	91	310
		ACAG	109	5193	1264	5511	1166
		ACAG	161	107	155	36	88
		ACAG	187	144	86	72	44
		ACAG	217	80	95	107	14
		ACAG	267	244	81	224	142
		ACAG	297	58	12	55	12
		ACAG	357	112	205	86	186
		ACAG	391	76	137	61	66
		ACAT	186	238	1070	747	539
		ACAT	267	48	112	68	238
		ACAT	344	114	248	260	110
		ACAT	348	244	416	588	271
		ACAT	358	79	58	175	80
		ACAT	368	96	239	73	138
		ACAT	371	164	158	71	176
		ACAT	374	106	94	59	88

		ACAT	384	179	161	182	262
		ACAT	393	1987	1391	1320	1708
		ACCA	142	902	1296	355	576
		ACCA	145	812	917	348	500
15	BAR1_34	ACCA	296	296	332	54	368
		ACCA	439	68	38	74	130
		ACCC	114	142	130	456	211
		ACCC	272	145	102	18	82
		ACCC	300	123	58	18	39
16	BAR1_35	ACCC	310	46	60	464	106
		ACCG	110	756	167	462	182
		ACCG	131	66	262	105	369
		ACCG	188	225	86	136	152
7	BAR1_37	ACCG	223	614	88	319	64
17	BAR1_38	ACCG	263	284	545	80	916
		ACCG	296	192	262	149	342
		ACCG	300	166	684	62	184
		ACCG	336	310	423	114	402
		ACCG	359	561	413	1192	160
		ACCG	482	96	69	41	40
		ACCT	249	65	170	192	102
		ACCT	435	45	29	66	28
		ACGA	115	118	375	1172	126
		ACGA	132	123	270	472	773
		ACGA	244	210	102	166	268
		ACGA	313	68	116	44	44
		ACGA	317	40	130	38	90
		ACGA	339	112	156	160	281
		ACGA	364	2572	1376	2916	444
		ACGA	397	56	98	136	278
		ACGA	400	64	93	177	288
		ACGC	77	1758	1870	2470	942
		ACGC	105	1679	2479	175	1328
		ACGC	111	940	1179	374	1132
		ACGC	138	212	170	554	220
		ACGC	142	208	226	460	112
8	BAR1_39	ACGC	203	75	262	37	46
		ACGC	238	100	116	27	48
		ACGC	268	111	76	198	171
		ACGC	302	218	239	228	94
		ACGG	191	727	1164	2414	1217
		ACGG	265	60	125	43	116
		ACGG	287	146	417	40	103
9	BAR1_7	ACGG	353	193	125	1043	144
		ACGG	356	105	125	573	68
		ACGG	367	37	28	22	52
		ACGG	385	86	174	40	66
		ACGG	423	86	56	32	66

		ACGG	426	71	36	31	50
		ACGG	447	100	53	546	78
1	BAR1_27	ACGG	458	88	1272	265	977
		ACGG	480	64	24	28	33
		ACGT	133	551	502	198	291
		ACGT	155	1833	806	1458	1641
		ACGT	190	1136	3666	966	1546
		ACGT	231	531	312	800	594
		ACGT	234	98	19	506	277
		ACGT	245	106	65	30	70
		ACGT	249	70	68	26	83
		ACGT	322	118	89	204	126
		ACGT	346	89	217	192	124
		ACTA	187	145	264	131	315
		ACTA	384	56	69	35	22
		ACTA	387	52	49	16	30
		ACTA	462	79	94	126	181
		ACTC	118	308	490	110	267
		ACTC	343	1585	384	2023	446
		ACTC	428	54	34	16	25
		ACTC	439	1278	1057	2796	1830
		ACTG	344	265	189	368	463
		ACTG	364	82	368	110	18
		ACTG	367	62	199	62	26
		ACTT	190	89	200	41	128
		ACTT	199	226	84	325	277
		ACTT	315	62	48	162	50
		ACTT	363	58	100	44	146
		ACTT	481	56	36	20	25
		AGAA	104	372	1354	232	749
		AGAA	197	1427	1236	246	1008
		AGAA	263	119	1005	342	174
		AGAA	323	105	38	160	86
		AGAA	399	250	560	184	427
		AGAA	466	78	216	303	317
		AGAC	267	42	32	103	58
		AGAC	291	92	44	115	64
		AGAC	337	46	15	42	28
		AGAC	345	107	148	53	45
		AGAG	297	48	68	122	49
		AGAG	341	60	20	16	38
		AGAG	486	35	40	58	25
		AGAG	498	27	56	23	34
		AGAT	80	190	280	76	216
		AGAT	244	268	359	84	242
		AGCA	296	510	618	1261	618
		AGCA	314	58	142	21	111
		AGCA	338	88	188	28	33

		AGCA	357	136	104	206	328
		AGCA	364	56	42	97	116
		AGCA	421	56	110	132	70
		AGCC	169	2672	2124	1106	2066
		AGCC	258	161	503	600	318
		AGCC	264	210	260	54	56
		AGCC	288	94	335	81	227
		AGCC	291	84	352	81	227
		AGCC	329	526	678	204	345
		AGCG	132	971	816	263	674
		AGCG	378	358	220	578	256
		AGCG	395	2155	2292	2480	2272
		AGCG	400	2336	2360	2675	3769
		AGCG	423	24	28	27	83
		AGCT	143	156	196	632	200
		AGCT	193	174	152	48	80
		AGCT	232	64	114	140	102
		AGCT	262	411	312	142	162
		AGCT	402	296	453	790	373
		AGCT	422	280	314	97	116
		AGGA	184	120	163	127	182
		AGGA	189	89	231	570	154
		AGGA	193	120	98	66	84
		AGGA	196	78	76	40	67
		AGGA	199	80	64	42	32
		AGGA	209	1812	2405	1794	1408
		AGGA	219	263	245	453	504
		AGGA	224	74	105	146	104
		AGGA	230	299	386	654	688
		AGGA	264	162	128	92	52
		AGGA	370	492	316	483	843
		AGGC	250	411	44	142	25
		AGGC	262	54	84	8	32
		AGGC	291	656	170	1059	320
		AGGC	296	373	66	650	195
		AGGC	343	38	43	16	43
		AGGC	382	32	242	28	30
		AGGC	438	30	54	94	38
		AGGG	158	184	171	67	108
		AGGG	202	340	320	77	54
		AGGG	218	563	1038	1292	1012
		AGGG	320	44	28	20	16
		AGGG	336	66	42	24	12
		AGGG	339	65	36	30	22
		AGGG	346	62	116	36	50
		AGGG	413	166	56	96	51
		AGGT	213	93	44	44	36
		AGGT	219	194	169	60	142

		AGGT	291	78	31	138	65
		AGGT	486	782	1528	1254	1343
		AGGT	492	359	600	652	611
		AGTA	287	62	90	63	24
		AGTA	308	80	31	19	60
		AGTA	358	72	44	84	154
		AGTA	384	159	117	26	115
		AGTA	453	74	59	61	24
10	BAR1_29	AGTC	96	1166	518	1652	886
		AGTC	266	30	101	25	62
		AGTC	283	124	91	44	28
		AGTC	314	142	149	58	116
		AGTC	342	44	28	58	24
		AGTC	373	116	105	105	34
		AGTC	438	90	60	294	136
		AGTG	103	164	122	108	66
		AGTG	273	64	70	53	34
		AGTG	276	52	58	36	34
18	BAR1_41	AGTG	434	208	182	299	20
		AGTG	438	118	101	86	22
		AGTT	204	76	181	285	92
		AGTT	246	104	124	142	54
		AGTT	294	96	64	34	82
		AGTT	453	67	112	76	50
		ATAA	238	42	34	82	44
		ATAA	259	693	2055	473	1526
		ATAA	316	140	140	332	168
		ATAA	322	212	289	550	319
		ATAA	462	153	238	508	678
		ATAC	222	260	196	178	58
		ATAC	267	106	40	218	142
		ATAC	293	48	68	26	75
		ATAC	315	107	278	426	331
		ATAG	206	60	103	22	16
		ATAG	246	84	174	142	251
		ATAG	267	118	62	78	142
		ATAG	488	52	30	27	23
		ATAT	162	296	1420	229	495
		ATAT	462	118	147	639	588
		ATAT	467	133	131	650	566
		ATCA	143	328	554	92	244
		ATCA	299	230	483	354	540
		ATCA	369	64	40	116	76
		ATCC	103	232	272	232	120
		ATCC	114	529	148	83	64
		ATCC	121	445	335	182	554
		ATCC	162	420	929	75	933
		ATCC	215	153	138	536	167

		ATCC	266	315	132	146	112
		ATCC	288	632	994	1050	1760
		ATCC	303	1000	744	403	667
		ATCC	380	48	146	46	58
		ATCC	438	296	72	348	282
		ATCG	442	63	67	33	73
		ATCG	452	46	102	28	46
		ATCG	464	1398	986	2000	1440
		ATCT	466	76	83	197	218
		ATGA	203	468	367	280	706
		ATGA	302	138	177	30	63
		ATGC	251	808	1180	197	1489
		ATGC	296	142	44	294	65
		ATGC	373	762	732	302	517
		ATGT	100	917	722	336	763
		ATGT	133	253	1142	149	248
		ATGT	305	136	246	88	124
		ATGT	322	122	154	362	117
		ATGT	361	158	110	56	78
		ATGT	427	92	95	52	42
		ATTA	119	54	29	43	80
		ATTA	210	50	58	28	82
		ATTA	223	166	144	33	62
		ATTA	303	64	130	28	72
		ATTA	454	48	91	113	20
		ATTA	474	74	36	56	146
		ATTA	480	51	29	16	34
		ATTC	125	132	110	380	206
		ATTC	171	54	52	84	25
		ATTC	191	321	246	142	252
		ATTC	235	176	294	2198	978
		ATTC	247	76	71	88	21
		ATTC	355	61	114	35	112
		ATTC	486	63	50	51	18
		ATTC	498	55	34	17	20
		ATTG	406	68	82	42	226
		ATTG	435	127	90	133	18
		ATTT	155	149	262	102	162
		ATTT	318	58	214	308	62
		ATTT	322	60	260	175	71
		ATTT	462	415	987	1280	2327
		ATTT	479	92	72	31	72
		ATTT	488	84	62	23	49
		CAAA	168	1088	1514	297	810
		CAAA	200	38	40	103	42
		CAAA	251	57	208	134	200
		CAAA	396	52	33	51	24
		CAAA	445	278	288	724	714

		CAAA	482	102	77	33	39
		CAAA	495	50	39	30	10
		CAAC	199	224	397	119	256
		CAAC	232	37	30	66	103
		CAAC	381	49	22	209	163
		CAAT	250	198	194	316	126
		CACA	117	132	351	280	88
		CACA	298	56	154	110	98
		CACA	301	64	183	124	140
		CACA	316	58	12	24	55
		CACA	436	190	466	288	225
		CACA	462	43	18	30	98
		CACG	105	560	486	1634	596
		CACG	138	522	188	301	490
		CACG	206	46	47	130	80
		CACG	218	284	220	90	91
		CACG	222	202	153	70	148
		CACG	250	53	138	20	31
		CACG	284	39	40	120	65
		CACG	374	45	148	66	394
		CACT	169	198	393	133	138
		CACT	189	62	178	100	59
2	BAR1_21	CAGA	106	3915	3668	4018	781
		CAGA	137	149	88	67	86
		CAGA	148	110	84	113	276
		CAGA	151	116	101	72	234
		CAGA	155	280	272	107	180
		CAGA	350	642	426	902	821
		CAGA	381	94	149	71	101
		CAGA	385	82	106	40	224
		CAGA	490	62	47	114	204
		CAGC	90	1503	1321	335	1134
		CAGC	118	1724	1641	456	3528
		CAGC	148	320	264	134	224
		CAGC	286	250	196	56	207
		CAGC	302	520	213	526	390
		CAGC	311	69	152	165	24
		CAGC	359	133	43	203	96
		CAGC	378	328	874	540	1112
		CAGC	416	79	140	158	38
		CAGG	438	48	78	40	30
		CAGT	215	134	66	23	30
		CATA	85	256	104	193	195
		CATA	151	844	628	228	962
		CATA	485	40	47	145	22
		CATC	155	1257	1450	1878	1380
		CATC	301	67	17	63	55
		CATC	369	254	196	504	280

		CATC	433	47	66	59	150
11	BAR1_49	CATC	450	762	701	644	80
		CATC	453	597	477	507	80
		CATG	223	2160	1518	882	1436
		CATG	247	144	54	62	135
		CATG	374	126	969	535	2024
		CATG	408	74	138	51	62
		CATG	423	248	330	152	82
		CATT	157	278	212	294	99
		CATT	205	52	307	78	428
		CATT	211	466	199	145	338
		CATT	222	988	279	1222	434
		CATT	225	1780	251	1916	212
		CATT	228	1703	80	1568	259
		CATT	235	694	80	959	88
		CATT	238	648	190	772	66
		CATT	242	354	434	194	190
		CATT	257	100	272	48	181
		CATT	264	120	235	50	123
		CATT	283	488	890	651	1436
		CATT	388	74	90	145	38
		CATT	451	1310	2609	3355	4230
		CCAA	155	1688	2056	330	1888
		CCAA	189	57	200	36	51
		CCAA	222	395	243	732	577
		CCAA	282	36	81	75	103
		CCAA	300	46	57	16	37
		CCAA	308	82	14	174	43
		CCAA	312	111	16	190	58
		CCAA	332	88	125	24	78
		CCAA	336	76	64	18	83
		CCAC	148	1560	616	2222	1609
		CCAC	199	2036	4496	2317	3961
		CCAC	227	96	55	172	178
		CCAC	336	242	69	314	244
		CCAC	341	234	96	275	241
		CCAC	351	33	122	62	342
		CCAC	357	25	88	55	247
		CCAC	459	184	607	176	515
		CCAC	462	101	501	85	278
		CCAG	186	153	264	98	162
		CCAG	246	200	452	169	122
		CCAG	311	26	160	18	164
		CCAG	318	28	66	30	34
		CCAG	453	110	146	254	112
		CCAT	96	274	84	403	164
		CCAT	206	947	153	1110	270
		CCAT	394	20	32	68	52

		CCCA	265	123	37	106	70
		CCCA	268	149	37	106	70
		CCCA	296	52	62	206	78
		CCCC	272	27	26	17	45
		CCCC	276	28	30	18	48
		CCCG	249	54	89	91	20
		CCCG	285	105	82	174	44
		CCCG	308	13	95	33	408
		CCCG	321	97	97	34	40
		CCCG	332	80	104	95	32
		CCCT	152	1048	2142	1216	5182
		CCCT	404	274	147	49	84
		CCGA	118	62	18	68	108
		CCGA	126	71	261	250	96
19	BAR1_50	CCGA	133	97	174	1432	229
		CCGA	143	105	32	761	70
		CCGA	191	447	765	283	329
		CCGA	274	64	16	28	68
		CCGA	310	159	348	236	52
		CCGA	317	94	104	124	38
		CCGA	363	34	14	26	51
		CCGA	429	36	35	47	75
		CCGC	290	216	208	220	51
		CCGG	249	238	582	326	58
		CCGT	93	193	397	136	390
		CCGT	130	844	1424	975	320
		CCGT	136	648	548	426	329
		CCGT	139	465	360	227	252
		CCGT	144	129	201	172	166
3	BAR1_11	CCGT	151	174	224	158	1004
		CCGT	163	1220	832	1060	979
		CCGT	169	375	168	896	298
		CCGT	176	184	200	370	220
		CCGT	210	41	22	48	24
		CCGT	234	1292	1045	388	908
		CCGT	443	722	1937	1231	846
		CCGT	475	50	18	36	28
		CCTA	110	582	315	144	363
		CCTA	193	914	386	1353	797
		CCTA	236	77	173	31	52
		CCTC	141	484	730	118	528
		CCTC	225	18	40	38	42
		CCTC	279	725	356	690	1074
		CCTC	301	20	28	95	60
		CCTC	305	38	26	114	14
		CCTC	319	135	176	54	150
		CCTC	335	4934	477	3218	5489
		CCTG	130	246	282	96	479

		CCTG	246	900	532	235	591
		CCTG	262	24	21	52	32
		CCTG	265	12	167	32	32
12	BAR1_52	CCTG	271	69	532	58	66
		CCTG	309	34	85	62	95
		CCTG	315	24	14	49	57
		CCTG	369	556	174	696	355
		CCTG	446	197	362	374	138
		CCTG	476	46	64	28	289
		CCTT	153	1438	488	1769	808
		CCTT	256	137	282	56	189
		CCTT	426	57	132	89	54
		CGAA	101	1705	2786	1123	3156
		CGAA	155	1664	1420	393	1224
		CGAA	239	611	910	160	263
		CGAA	385	42	24	60	66
		CGAC	96	5674	6369	1352	4262
		CGAC	176	502	193	868	552
		CGAC	232	68	30	94	48
		CGAC	419	124	100	34	42
		CGAG	122	534	237	208	225
		CGAG	161	577	1112	443	451
		CGAT	94	2152	1062	185	2826
		CGAT	251	581	360	849	389
		CGAT	305	54	377	140	127
		CGCA	114	1016	975	2760	1723
		CGCA	162	1036	1256	366	1599
		CGCA	267	26	17	60	86
		CGCA	289	45	56	20	33
		CGCA	363	5780	4633	2042	4229
		CGCA	411	75	312	112	391
		CGCA	472	44	126	70	158
		CGCG	122	130	177	484	150
		CGCG	131	702	285	394	1649
		CGCG	161	150	374	166	317
		CGCG	263	266	201	100	316
		CGCG	308	20	149	90	248
		CGCG	334	149	208	135	97
		CGCG	386	40	24	158	64
		CGCT	109	160	352	55	161
		CGCT	152	556	872	531	2286
		CGCT	253	52	76	26	72
		CGCT	283	60	78	26	36
		CGCT	298	44	64	36	22
		CGCT	337	27	48	126	34
		CGCT	354	274	179	206	759
		CGCT	367	225	114	100	432
		CGCT	373	124	138	47	106

		CGCT	401	58	33	33	20
		CGGA	108	773	601	1997	186
		CGGA	227	659	426	288	339
		CGGA	240	112	216	76	126
		CGGA	309	452	1091	1368	932
		CGGA	425	57	852	106	608
		CGGA	451	97	43	44	44
		CGGC	105	343	242	233	150
		CGGC	376	20	59	27	40
		CGGG	218	364	700	263	302
		CGGG	341	457	359	80	296
		CGGG	358	109	961	138	296
		CGGT	110	284	350	758	650
		CGGT	127	1286	778	572	718
		CGGT	138	246	298	606	298
		CGGT	261	69	145	39	250
		CGGT	279	262	290	484	215
		CGGT	392	282	144	285	118
		CGGT	425	50	118	89	88
		CGTA	241	73	90	36	80
		CGTA	245	21	210	76	68
		CGTA	264	3104	2038	1300	1922
		CGTA	325	194	191	98	388
		CGTC	310	80	126	52	64
		CGTC	335	764	150	746	1110
		CGTC	341	74	26	155	55
		CGTC	344	29	9	57	22
		CGTG	160	175	762	616	312
		CGTG	235	238	93	41	94
		CGTG	238	148	84	37	79
		CGTG	244	68	59	26	43
		CGTG	252	62	28	17	283
		CGTG	308	15	17	30	38
		CGTT	225	460	50	420	132
		CGTT	256	114	244	54	282
		CGTT	423	50	61	48	18
		CGTT	450	122	214	523	791
		CTAA	80	280	304	286	123
		CTAA	155	551	524	108	738
		CTAA	383	66	26	42	59
		CTAC	86	92	258	301	98
		CTAC	186	50	514	667	262
		CTAC	189	99	346	526	204
		CTAG	133	198	78	141	128
		CTAG	143	1131	453	486	635
		CTAG	163	344	140	59	152
		CTAG	366	24	168	34	66
		CTAG	371	40	141	22	45

		CTAG	467	87	52	44	106
14	BAR1_22	CTAT	95	1489	1706	439	1496
		CTAT	293	49	28	32	16
		CTCC	95	26	76	62	72
		CTCC	308	1146	1107	1735	124
		CTCC	312	790	659	1244	91
		CTCC	348	105	734	277	1832
		CTCC	484	71	412	112	784
		CTCG	113	122	86	214	75
		CTCG	253	312	139	386	102
4	BAR1_28	CTCG	310	90	2989	184	3740
		CTCG	314	245	1800	252	2562
		CTCG	328	46	86	73	246
		CTCG	346	50	8	44	85
		CTCG	356	81	22	22	17
		CTCG	474	42	34	22	66
		CTCT	380	128	306	220	174
		CTCT	391	68	19	86	88
		CTCT	406	286	994	474	554
		CTGC	153	233	406	48	148
		CTGC	159	547	903	204	544
		CTGC	307	160	145	216	51
		CTGC	365	35	17	52	48
		CTGG	115	849	2168	986	172
		CTGG	237	238	362	102	72
		CTGG	246	128	316	226	61
		CTGG	255	158	204	97	32
		CTGG	284	104	390	129	3718
		CTGG	291	529	312	308	2050
		CTGG	302	89	60	138	136
		CTGT	132	1510	1468	546	1886
		CTGT	398	46	91	136	83
		CTGT	409	30	46	70	47
		CTTA	243	92	46	25	27
		CTTA	299	26	42	91	42
		CTTC	85	897	1385	232	353
		CTTC	88	801	1180	222	1820
		CTTC	110	944	604	1328	1473
		CTTC	167	332	724	784	243
		CTTC	199	145	98	76	1512
		CTTC	335	842	96	542	397
		CTTG	162	306	492	171	320
		CTTG	258	1628	451	1278	403
		CTTG	331	273	121	80	116
		CTTG	467	212	116	113	453
		CTTT	95	448	245	340	846
		CTTT	238	36	52	16	32
		CTTT	274	144	94	275	276

		CTTT	364	48	67	14	59
		CTTT	411	100	82	326	43
		CTTT	455	31	68	82	50
		GAAA	105	450	293	199	267
		GAAA	150	171	177	72	58
		GAAA	360	66	126	128	52
		GAAC	139	775	646	204	696
		GAAC	164	74	119	72	173
		GAAC	279	98	74	52	34
		GAAG	100	798	277	658	418
		GAAG	140	198	1286	168	628
		GAAG	182	913	604	1730	1011
		GAAG	213	475	552	226	238
		GAAG	221	264	144	102	207
		GAAG	421	50	20	73	24
		GAAT	277	230	124	38	134
		GACC	135	72	106	48	126
		GACC	160	414	199	647	74
		GACC	164	247	28	424	77
		GACC	184	190	518	216	309
		GACG	100	227	292	41	268
		GACG	167	434	1070	364	241
		GACG	204	77	95	224	204
		GACG	333	77	114	24	83
		GACG	451	65	51	24	24
		GACG	467	147	166	358	445
		GACT	344	43	170	147	121
		GAGA	257	222	618	122	212
		GAGC	133	117	252	256	92
		GAGC	141	224	356	68	166
		GAGC	178	197	214	338	272
		GAGC	182	98	138	232	214
		GAGC	187	73	78	86	116
		GAGC	197	294	399	174	406
		GAGC	207	992	994	977	853
		GAGC	215	107	55	182	202
		GAGC	220	106	38	36	68
		GAGG	312	67	117	66	28
		GAGG	318	44	909	22	32
		GAGG	321	52	380	52	15
		GAGG	436	42	64	26	45
		GAGG	483	56	60	20	46
		GAGT	195	330	780	260	222
		GAGT	214	2773	2669	2236	3638
		GAGT	427	1062	1078	1216	1008
		GATA	250	92	44	108	44
		GATC	162	86	76	223	70
		GATC	179	96	356	51	318

		GATC	263	52	50	26	79
		GATC	339	132	155	27	62
		GATC	345	67	66	32	50
		GATT	194	66	140	50	58
		GATT	197	52	132	74	71
		GATT	337	24	70	32	36
		GCAA	109	280	333	90	84
		GCAA	162	138	147	52	111
		GCAA	186	174	220	34	64
		GCAA	249	154	116	272	51
		GCAA	252	133	76	192	33
		GCAA	343	78	27	76	34
		GCAA	379	38	101	92	98
		GCAC	91	227	176	528	240
		GCAC	189	761	311	158	40
		GCAC	335	36	109	34	56
		GCAG	111	1306	509	836	824
		GCAG	209	538	502	747	637
		GCAG	244	64	56	218	140
		GCAG	262	61	133	29	44
		GCAG	317	36	124	46	45
		GCAT	165	226	236	172	36
		GCCA	131	93	110	24	83
		GCCA	165	261	381	123	382
		GCCA	202	92	101	30	57
		GCCA	333	68	82	331	17
		GCCC	143	268	566	462	892
		GCCC	228	280	116	202	268
		GCCC	232	184	34	75	127
		GCCC	307	52	138	112	85
		GCCC	356	81	124	204	62
		GCCC	390	84	109	36	78
		GCCG	104	227	142	58	116
		GCGA	165	108	104	21	174
		GCGA	288	25	28	127	96
		GCGA	292	60	126	37	64
		GCGA	300	39	59	19	19
		GCGA	317	85	32	78	84
		GCGA	391	64	121	51	34
		GCGA	397	56	104	40	32
		GCGC	88	1065	936	1143	293
		GCGC	315	242	221	74	132
		GCGG	262	67	140	212	204
		GCGT	327	72	118	35	62
		GCTA	173	108	80	137	52
		GCTA	204	356	587	180	721
		GCTA	284	153	57	363	221
		GCTA	346	82	40	56	30

		GCTC	91	952	1054	371	800
		GCTC	112	1226	760	2316	1004
		GCTC	125	3047	5945	1256	6320
13	BAR1_8	GCTC	140	273	479	1316	236
5	BAR1_24	GCTC	167	142	2227	1510	1923
		GCTC	197	185	256	109	88
		GCTC	213	204	400	142	208
		GCTC	338	140	242	20	166
		GCTC	342	213	271	92	228
		GCTC	386	1234	1722	1154	415
		GCTG	266	78	85	65	25
		GCTG	271	117	108	234	126
		GCTG	274	74	120	166	112
		GCTG	277	113	70	71	83
		GCTG	289	360	456	122	164
		GCTG	294	149	182	32	239
		GCTG	299	51	292	12	240
		GCTG	302	48	234	26	144
		GCTG	305	59	69	26	52
		GCTG	310	84	76	58	38
		GCTG	313	194	340	184	140
		GCTG	345	221	99	640	60
		GCTG	354	1223	1024	748	691
		GCTG	362	1079	866	681	1392
		GCTG	371	201	76	412	218
		GCTG	409	42	68	25	69
		GCTG	427	40	54	15	30
		GCTT	178	344	326	84	270
		GCTT	245	68	166	114	99
		GGAA	156	69	136	36	45
		GGAA	159	82	127	42	58
		GGAA	377	64	58	34	23
		GGAC	90	4102	4249	1818	3616
		GGAC	127	260	452	575	196
		GGAC	199	226	435	28	144
		GGAC	216	702	133	510	418
		GGAC	219	668	60	458	408
		GGAC	242	72	74	257	117
		GGAC	250	106	188	176	60
		GGAC	254	66	164	150	26
		GGAC	308	157	98	46	128
		GGAC	371	22	44	222	36
		GGAG	154	46	536	234	118
		GGAG	176	1115	362	512	1340
		GGAG	193	193	234	96	194
		GGAG	322	25	26	58	54
		GGAG	360	168	289	807	159
		GGAT	107	712	590	434	196

		GGAT	123	145	529	320	1274
		GGAT	143	196	490	565	256
		GGAT	153	510	488	188	564
		GGAT	182	1273	1571	1498	414
		GGAT	209	340	466	425	171
		GGAT	258	86	58	179	202
		GGAT	314	37	11	58	53
		GGCA	304	514	653	225	517
		GGCA	341	33	66	50	110
		GGCC	172	164	52	100	222
		GGCC	206	119	262	420	184
		GGCC	326	74	36	46	130
		GGCC	349	87	372	70	72
		GGCC	429	38	58	20	37
		GGCG	107	196	122	90	646
		GGCG	218	456	584	162	570
		GGCT	91	1380	590	1028	262
		GGCT	164	155	228	46	220
		GGCT	171	16	78	64	49
		GGGA	164	120	106	257	205
		GGGA	258	1128	4752	1738	2970
		GGGA	466	844	590	1807	2284
		GGGC	101	1152	742	338	1296
		GGGC	189	30	14	51	18
		GGGC	194	142	214	72	220
		GGGC	246	68	78	34	74
20	BAR1_62	GGGC	338	16	540	54	14
		GGGC	379	50	12	18	36
		GGGC	424	30	113	37	34
		GGGC	449	29	52	22	12
		GGGC	452	52	58	23	17
		GGGC	480	14	40	11	8
		GGGC	484	40	71	20	17
		GGGG	95	373	148	1770	804
		GGGG	112	503	424	148	379
		GGGG	121	474	291	150	255
		GGGG	140	36	191	89	545
		GGGG	241	186	124	182	70
		GGGG	486	415	490	110	613
		GGGT	161	64	140	140	106
		GGGT	314	216	56	45	44
		GGTA	277	74	56	164	72
		GGTA	371	277	68	652	190
		GGTA	432	39	120	52	56
		GGTC	106	96	322	100	186
		GGTC	170	282	176	108	76
		GGTC	222	207	289	584	410
		GGTC	310	40	79	43	107

		GGTG	81	632	1636	802	2008
		GGTG	163	508	274	581	149
		GGTG	177	1098	956	1270	394
		GGTG	191	4182	6469	1196	4235
		GGTG	246	120	328	149	502
		GGTG	280	636	328	324	182
		GGTG	340	1796	3596	4212	560
		GGTG	343	4319	6913	6032	1624
		GGTG	392	70	46	22	26
		GGTG	412	50	44	14	27
		GGTT	77	774	533	162	684
		GGTT	251	89	60	202	27
		GGTT	338	95	289	75	82
		GGTT	377	111	186	68	162
		GTAA	113	83	126	104	387
		GTAA	204	57	44	14	81
		GTAA	269	44	24	64	58
		GTAA	314	384	300	159	270
		GTAA	362	248	232	362	146
		GTAA	385	44	24	73	48
		GTAC	207	108	86	62	34
		GTAC	237	374	633	169	664
		GTAG	130	146	116	104	46
		GTAG	133	28	86	176	110
		GTAG	181	806	228	1230	1252
		GTAG	197	446	413	283	212
		GTAG	216	292	908	619	727
		GTAT	170	323	746	342	1280
		GTAT	221	154	200	728	356
		GTAT	269	298	416	339	618
		GTAT	343	128	68	227	114
		GTAT	484	79	32	19	44
		GTCA	106	530	535	109	318
		GTCA	227	60	29	42	66
		GTCA	339	45	78	29	115
		GTCA	342	32	89	38	159
		GTCA	455	53	52	98	122
		GTCC	124	654	610	155	392
		GTCC	301	24	51	25	21
		GTCC	336	75	27	46	22
		GTCG	137	114	192	348	136
		GTCG	161	248	472	823	286
		GTCG	204	256	177	63	372
		GTCG	355	64	72	174	100
		GTCG	452	98	76	37	129
		GTCT	200	52	114	46	52
		GTCT	347	94	68	37	48
		GTGA	110	210	304	99	195

		GTGA	122	180	291	217	132
		GTGA	193	398	504	335	452
		GTGA	197	346	383	330	310
		GTGA	207	924	614	1166	799
		GTGA	212	495	274	551	492
		GTGA	222	282	564	204	156
		GTGA	225	282	503	164	165
		GTGA	233	71	39	38	40
		GTGA	291	92	134	28	111
		GTGC	84	339	142	567	423
		GTGC	95	1898	1974	2604	1014
		GTGC	100	732	1775	3712	2654
		GTGC	117	422	832	356	793
		GTGC	223	71	74	79	152
		GTGG	111	4247	4086	1669	2826
		GTGG	163	254	676	590	501
		GTGG	293	79	34	78	84
		GTGT	133	144	104	178	270
		GTGT	170	180	234	54	60
		GTGT	192	5606	5318	5272	5324
		GTGT	320	108	88	40	66
		GTTA	164	985	935	136	818
		GTTA	455	154	57	366	58
		GTTC	111	301	136	437	368
		GTTC	167	66	546	176	375
		GTTC	362	633	748	582	238
		GTTC	410	83	132	31	82
		GTTG	224	57	36	88	64
		GTTG	299	136	328	276	296
		GTTG	318	78	50	29	40
		GTTG	401	48	27	49	68
		GTTT	98	263	285	270	52
		GTTT	223	20	33	28	50
		TAAA	78	538	990	246	711
		TAAA	118	49	135	78	58
		TAAA	481	48	53	124	97
		TAAC	490	76	42	60	10
		TAAG	201	56	26	62	82
		TAAG	488	62	20	10	7
		TAAG	491	47	26	14	14
		TAAT	146	259	796	378	624
		TACA	156	266	80	214	89
		TACA	161	206	258	76	232
		TACA	256	34	60	13	25
		TACC	139	1104	680	1550	972
21	BAR1_20	TACC	176	1590	1177	270	1116
		TACC	185	808	2248	1378	942
		TACC	222	676	272	453	807

		TACC	249	450	82	300	863
		TACC	262	1237	931	458	543
		TACC	306	196	82	190	116
		TACC	489	39	39	26	18
		TACG	208	131	27	60	95
		TACG	328	34	62	26	40
		TACG	462	2076	3834	5031	6352
		TACG	467	2005	3507	4998	6147
		TACG	494	46	31	10	10
		TACG	498	54	29	18	0
		TACT	162	40	337	261	50
		TACT	203	70	138	56	26
		TACT	492	48	36	20	13
		TAGA	215	1488	37	221	239
22	BAR1_67	TAGC	119	372	454	117	356
		TAGC	311	99	18	178	24
		TAGC	421	54	32	28	20
		TAGC	462	1221	1442	1464	2680
		TAGC	467	1185	1481	1952	2532
		TAGC	485	62	28	36	8
		TAGG	174	78	208	158	169
		TAGG	259	80	125	114	38
		TAGG	381	68	18	41	32
		TAGG	395	58	67	88	36
		TAGG	428	70	35	20	26
		TAGG	449	93	36	118	57
		TAGG	462	1179	1632	2440	3028
		TAGG	467	1149	1555	2420	2878
		TAGG	483	92	27	20	19
		TAGT	207	380	900	450	173
		TAGT	496	52	24	10	15
		TATA	214	535	70	64	62
		TATA	357	217	337	74	336
		TATA	481	36	46	114	91
		TATC	176	214	101	40	76
		TATC	346	28	20	52	44
		TATC	361	45	104	14	33
		TATC	385	54	18	33	48
		TATC	462	3024	4121	4736	6736
		TATC	467	2944	3871	4796	6604
		TATC	492	78	44	30	26
		TATC	495	72	61	6	21
		TATG	180	82	199	197	756
		TATG	183	128	260	385	340
		TATG	245	178	68	190	101
		TATG	310	179	362	44	52
		TATG	499	83	28	12	12
		TATT	319	1148	106	201	243

		TATT	481	56	65	133	115
		TCAA	155	505	448	134	584
		TCAA	158	505	310	134	615
		TCAC	196	80	398	347	372
		TCAC	228	64	62	144	88
		TCAC	249	484	670	750	239
		TCAC	290	2690	682	2508	1937
		TCAC	352	52	170	32	176
		TCAG	109	238	162	493	238
		TCAG	134	1179	239	1268	149
		TCAG	140	408	199	336	95
		TCAG	154	1959	3091	1047	3389
		TCAG	194	79	180	115	77
		TCAG	400	44	26	70	58
		TCAT	89	292	652	900	1040
		TCAT	256	342	193	272	72
		TCAT	262	1170	1587	202	824
		TCAT	408	514	842	227	506
		TCCA	196	1388	2316	895	2121
		TCCA	249	238	128	442	293
		TCCA	295	468	668	257	672
		TCCA	299	256	404	93	220
		TCCC	118	386	280	40	286
		TCCC	132	258	148	787	105
		TCCC	135	218	83	682	92
		TCCC	138	218	85	733	69
		TCCC	226	130	222	57	36
		TCCC	250	52	76	22	39
		TCCC	451	260	702	280	865
		TCCC	467	106	74	146	188
		TCCC	486	56	52	37	18
		TCCG	116	388	113	554	959
		TCCG	126	522	292	218	415
		TCCG	161	222	66	15	136
		TCCG	207	194	147	46	146
		TCCG	332	40	108	141	115
		TCCG	392	212	1027	42	346
		TCCT	152	250	362	260	706
		TCGA	126	62	93	276	56
		TCGA	133	94	65	945	108
		TCGA	308	153	130	104	53
		TCGA	446	164	163	322	110
		TCGA	484	56	22	6	0
		TCGC	79	491	192	702	223
		TCGC	111	178	54	198	301
		TCGC	198	210	104	86	96
		TCGC	201	200	120	86	62
		TCGG	139	168	100	40	309

		TCGG	208	1560	916	560	900
		TCGG	283	143	396	144	176
		TCGG	292	92	42	154	46
		TCGT	130	332	3136	848	336
		TCGT	146	84	95	57	152
		TCGT	251	128	330	197	231
		TCGT	480	108	10	46	42
		TCTA	165	50	82	26	96
		TCTA	325	4516	4616	3200	112
		TCTA	328	2725	2272	1724	117
		TCTA	331	1012	618	611	85
		TCTA	350	68	100	159	1367
		TCTA	428	78	496	60	218
		TCTC	220	64	62	35	80
		TCTC	335	539	232	286	1121
		TCTC	349	159	107	30	52
		TCTC	404	124	194	56	88
		TCTC	414	49	64	48	23
		TCTC	427	564	306	676	354
		TCTC	453	81	46	18	14
		TCTG	154	251	522	113	439
		TCTG	332	56	31	21	50
		TCTT	211	228	312	120	114
		TCTT	214	210	225	117	84
		TCTT	486	42	12	24	10
		TGAA	139	416	210	568	222
		TGAA	168	590	846	235	624
		TGAA	298	206	196	536	274
		TGAA	391	498	223	388	752
		TGAA	446	269	174	392	142
		TGAA	487	73	77	13	24
		TGAG	141	83	60	170	56
		TGAG	357	20	49	35	84
		TGAG	416	49	34	80	34
		TGAG	479	45	14	26	12
		TGAT	143	112	58	30	90
		TGAT	194	338	109	117	64
		TGAT	238	38	64	102	99
		TGAT	307	497	522	160	419
		TGAT	391	1506	1218	860	1953
		TGAT	485	64	17	12	19
		TGCA	111	379	263	697	420
		TGCA	187	902	1450	876	630
23	BAR1_69	TGCA	213	89	723	48	872
		TGCA	349	72	48	16	41
		TGCC	145	60	294	112	207
		TGCC	164	136	161	27	98
		TGCC	306	340	275	464	86

		TGCG	230	136	812	656	151
		TGCG	484	102	34	29	48
		TGCT	82	938	634	589	436
		TGCT	87	1019	1357	468	1145
		TGCT	194	740	502	288	672
		TGCT	203	166	759	36	59
		TGCT	281	74	19	24	61
		TGCT	307	422	872	444	84
		TGCT	310	288	624	303	79
		TGCT	483	52	54	26	20
		TGGA	174	1354	499	2090	1480
		TGGA	186	96	928	194	68
		TGGA	308	48	138	170	193
		TGGC	97	100	236	106	95
		TGGC	238	242	406	92	186
		TGGC	241	162	285	26	109
		TGGC	247	124	122	45	118
		TGGC	288	136	51	132	104
		TGGC	344	105	286	242	284
		TGGC	353	253	556	146	381
		TGGT	299	382	171	75	141
		TGTA	99	196	186	60	114
		TGTA	103	140	180	73	184
		TGTA	212	31	166	44	198
		TGTA	222	160	48	146	120
		TGTA	427	42	84	29	39
		TGTA	482	58	106	68	22
		TGTC	135	397	870	232	756
		TGTC	185	264	836	246	326
		TGTC	231	54	98	262	339
		TGTC	247	2474	1326	1938	3760
		TGTC	282	242	224	294	67
		TGTC	295	73	173	73	92
		TGTC	318	98	100	277	112
		TGTC	369	526	509	158	196
		TGTC	386	89	96	179	234
		TGTC	395	427	984	1567	1374
		TGTC	421	28	18	62	62
		TGTC	448	158	273	88	92
		TGTG	287	60	32	98	139
		TGTG	353	30	12	53	33
		TGTT	243	657	804	256	536
		TGTT	293	100	56	56	338
		TGTT	344	298	213	76	244
		TGTT	488	95	51	38	14
		TTAA	147	478	672	316	78
		TTAC	154	748	950	300	780
		TTAC	482	83	32	42	8

		TTAG	122	111	135	277	164
		TTAG	238	76	172	132	200
		TTAG	491	52	38	18	15
		TTAG	494	50	38	7	15
		TTAT	169	42	72	110	72
		TTAT	184	31	20	42	62
		TTAT	482	44	120	108	49
		TTCA	213	23	260	54	164
		TTCA	218	616	175	295	430
		TTCA	494	45	11	32	16
		TTCC	419	103	115	39	65
		TTCG	179	466	489	125	223
		TTCG	238	52	91	231	98
		TTCG	308	154	743	90	334
		TTCG	311	108	452	60	254
		TTCG	478	56	10	56	52
		TTGA	144	105	100	99	122
		TTGA	213	24	74	70	108
		TTGA	393	60	24	47	66
		TTGC	284	108	42	116	30
		TTGG	143	75	15	219	66
		TTGG	172	14	32	93	35
		TTGG	238	418	150	164	428
		TTGT	238	406	148	224	436
		TTGT	257	63	195	86	221
		TTGT	442	330	233	362	166
		TTTA	100	2088	2058	540	1588
		TTTC	243	162	187	77	103
		TTTG	172	70	44	48	26
		TTTG	312	66	82	54	21
		TTTT	327	993	284	341	98
		TTTT	356	132	472	248	66
		TTTT	446	515	458	960	343
		TTTT	451	456	476	601	243
		TTTT	491	40	36	14	53

TABLE 3

Seq ID	Clone ID	Digital Address (MspI)	Database Match (Accession #)	% Homology	Nucleotide homology	
					DST nucleotide range (bp#)	Database nucleotide range (bp#)
9	BAR1_7	ACGG 353	EST uj37f10.x1 Sugano mouse kidney mkia Mus musculus cDNA clone IMAGE:1922155 3' similar to TR:Q14120 Q14120 DBP-5 NUCLEAR PROTEIN (AI315677)	98%	1 – 298	84 – 381
13	BAR1_8	GCTC 140	EST UI-M-BH0-ajz-b-03-0-UI.s1 NIH_BMAP_M_S1 Mus musculus cDNA clone UI-M-BH0-ajz-b-03-0-UI 3', mRNA sequence (AI854173.1)	97%	1 – 89	13 – 101
3	BAR1_11	CCGT 151	Mus musculus soluble lectin (Mac-2) gene (L08649)	98%	1 – 95	6785 – 6879
6	BAR1_16	AACG 119	EST va36h12.r1 Soares mouse 3NME12 5 Mus musculus cDNA clone 733511 5' (AA259694)	97%	1 – 68	299 – 366
21	BAR1_20	TACC 176	Mouse (clone RAG11) T-cell receptor rearranged gamma-chain mRNA (M34970)	99%	1 – 124	71 – 194
2	BAR1_21	CAGA 106	EST vu32a04.r1 Stratagene mouse Tcell 937311 Mus musculus cDNA clone 1193070 5' (AA718318)	100%	1 – 51	68 – 118
14	BAR1_22	CTAT 95	EST mw38e07.r1 Soares mouse 3NME12 5 Mus musculus cDNA clone 672996 5' (AA231627)	100%	1 – 42	150 – 191
5	BAR1_24	GCTC 167	Mouse interleukin-4 receptor (secreted form) mRNA (M27960)	99%	1 – 107	3587 – 3693
1	BAR1_27	ACGG 458	Mouse granzyme C serine esterase mRNA (M18459)	99%	1 – 363	532 – 894
4	BAR1_28	CTCG 310	Mouse mRNA for peripheral-type benzodiazepine receptor (D21207)	98%	1 – 251	561 – 810

EST = Expressed Sequence Tag, N/A = Not Applicable

TABLE 3 (continued)

Seq ID	Clone ID	Digital Address (MspI)	Database Match (Accession #)	% Homology	Nucleotide homology	
					DST nucleotide range (bp#)	Database nucleotide range (bp#)
10	BAR1_29	AGTC 96	Mus musculus small intestine C57BL/6J adult Mus musculus cDNA clone 2010100N16, mRNA sequence (AV065690.1)	94%	1 – 37	86 – 122
15	BAR1_34	ACCA 296	Mus musculus K-ras type A mRNA, 3' untranslated sequence (U76425)	96%	9 – 246	345 – 580
16	BAR1_35	ACCC 310	Mus musculus annexin XI (Anx11) mRNA, complete cds. (U65986)	98%	1 – 258	1482 – 1739
7	BAR1_37	ACCG 223	EST ve84h05.r1 Soares mouse NbMH Mus musculus cDNA clone IMAGE:832953 5' (AA423053)	100%	1 – 166	146 – 311
17	BAR1_38	ACCG 263	Mus musculus chimeric 16S ribosomal RNA, complete sequence; mitochondrial gene for nuclear product (AF089815.1)	99%	3 – 209	1046 – 1252
8	BAR1_39	ACGC 203	EST vt24h03.x1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone IMAGE:1164053 3' similar to TR:O42573 O42573 NEDD4 PROTEIN (AI591551.1)	100% 97%	37 – 149 1 – 39	1 – 113 287 – 325
18	BAR1_41	AGTG 434	Mus musculus COP9 (constitutive, photomorphogenic), subunit (NM_0102001.1)	99%	1 – 385	1215 – 1599
11	BAR1_49	CATC 450	EST vw57f05.x1 Soares mouse mammary gland NMLMG Mus musculus cDNA clone IMAGE: 1247937 (AI461717)	98%	1 – 394	5 – 398
19	BAR1_50	CCGA 133	Mouse 28S ribosomal RNA (X00525)	94% 100%	1 – 35 29 – 81	3209 – 3244 3322 – 3374

TABLE 3 (continued)

Seq ID	Clone ID	Digital Address (MspI)	Database Match (Accession #)	% Homology	Nucleotide homology	
					DST nucleotide range (bp#)	Database nucleotide range (bp#)
12	BAR1_52	CCTG 271	EST vb18g06.x1 Soares mouse 3NbMS Mus musculus cDNA clone IMAGE:749338 3' similar to TR:O14862 O14862 INTERFERON-INDUCIBLE PROTEIN (AI661692.1)	98%	1 – 209	1 – 209
20	BAR1_62	GGGC 338	Mus musculus p162 protein mRNA, complete cds (U14172.1)	99% 88%	1 – 281 40 – 75	3262 – 3542 3088 – 3123
22	BAR1_67	TAGC 119	EST mm86f09.y1 Stratagene mouse embryonic carcinomaRA (#937318) Mus musculus cDNA clone IMAGE:535337 5' similar to SW:YAQ6_SCHPO Q10106 HYPOTHETICAL 29.7 KD PROTEIN C18G6.06 IN CHROMOSOME I; mRNA sequence (AI325714)	100%	5 – 68	55 – 118
23	BAR1_69	TGCA 213	Mus musculus advillin (Advil-pending), mRNA (NM_009635.1)	100%	1 – 159	269 – 427

TABLE 4

Seq ID	Clone ID	Digital Address (MspI)	Database Match (Accession #)	Extended Primer	Primer Seq ID
9	BAR1_7	ACGG 353	EST uj37f10.x1 Sugano mouse kidney mkia Mus musculus cDNA clone IMAGE:1922155 3' similar to TR:Q14120 Q14120 DBP-5 NUCLEAR PROTEIN (AI315677)	GAT CGA ATC CGG ACG GTG TAC CCC GAG GAT	30
13	BAR1_8	GCTC 140	EST UI-M-BH0-ajz-b-03-0-UI.s1 NIH_BMAP_M_S1 Mus musculus cDNA clone UI-M-BH0-ajz-b-03-0-UI 3', mRNA sequence (AI854173.1)	GAT CGA ATC CGG GCT CTT TGC ACA GTT CTA	31
3	BAR1_11	CCGT 151	Mus musculus soluble lectin (Mac-2) gene (L08649)	GAT CGA ATC CGG CCG TGT GTG CCT TAG GAG	32
6	BAR1_16	AACG 119	EST va36h12.r1 Soares mouse 3NME12 5 Mus musculus cDNA clone 733511 5' (AA259694)	GAT CGA ATC CGG AAC GGG GGA TGA TGG GGG	33
21	BAR1_20	TACC 176	Mouse (clone RAG11) T-cell receptor rearranged gamma-chain mRNA (M34970)	GAT CGA ATC CGG TAC CAA TGT ATA GCT GTG	34
2	BAR1_21	CAGA 106	EST vu32a04.r1 Stratagene mouse Tcell 937311 Mus musculus cDNA clone 1193070 5' (AA718318)	GAT CGA ATC CGG CAG AGG CTC CTG GCC TGC	35
14	BAR1_22	CTAT 95	EST mw38e07.r1 Soares mouse 3NME12 5 Mus musculus cDNA clone 672996 5' (AA231627)	GAT CGA ATC CGG CTA TCA GTG CTA ACA GAA	36
5	BAR1_24	GCTC 167	Mouse interleukin-4 receptor (secreted form) mRNA (M27960)	GAT CGA ATC CGG GCT CAG CAC TGT CTG CTC	37
1	BAR1_27	ACGG 458	Mouse granzyme C serine esterase mRNA (M18459)	GAT CGA ATC CGG ACG GGG AAT TCC CAA AAA	38
4	BAR1_28	CTCG 310	Mouse mRNA for peripheral-type benzodiazepine receptor (D21207)	GAT CGA ATC CGG CTC GCA GAG TGA AGG CAC	39
10	BAR1_29	AGTC 96	Mus musculus small intestine C57BL/6J adult Mus musculus cDNA clone 2010100N16, mRNA sequence (AV065690.1)	GAT CGA ATC CGG AGT CCC CTC CTC TGA ATT	40

EST = Expressed Sequence Tag

Table 4 (continued)

Seq ID	Clone ID	Digital Address (MspI)	Database Match (Accession #)	Extended Primer	Primer Seq ID
15	BAR1_34	ACCA 296	Mus musculus K-ras type A mRNA, 3' untranslated sequence (U76425)	GAT CGA ATC CGG ACC ATC CCT GCT CTG TGT	41
16	BAR1_35	ACCC 310	Mus musculus annexin XI (Anx11) mRNA, complete cds. (U65986)	GAT CGA ATC CGG ACC CTG ATT CGC ATC ATG	42
7	BAR1_37	ACCG 223	EST ve84h05.r1 Soares mouse NbMH Mus musculus cDNA clone IMAGE:832953 5' (AA423053)	GAT CGA ATC CGG ACC GCC ATG CCT GGT GAG	43
17	BAR1_38	ACCG 263	Mus musculus chimeric 16S ribosomal RNA, complete sequence; mitochondrial gene for nuclear product (AF089815.1)	GAT CGA ATC CGG ACC GTG CAA AGG TAG CAT	44
8	BAR1_39	ACGC 203	EST vt24h03.x1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone IMAGE:1164053 3' similar to TR:O42573 O42573 NEDD4 PROTEIN (AI591551.1)	GAT CGA ATC CGG ACG CTG GAG CCT TCC TTA	45
18	BAR1_41	AGTG 434	Mus musculus COP9 (constitutive, photomorphogenic), subunit (NM_0102001.1)	GAT CGA ATC CGG AGT GGA CAG CAC AAG CCA	46
11	BAR1_49	CATC 450	EST vw57f05.x1 Soares mouse mammary gland NMLMG Mus musculus cDNA clone IMAGE: 1247937 (AI461717)	GAT CGA ATC CGG CAT CGT CCT TGT TAA GCA	47
19	BAR1_50	CCGA 133	Mouse 28S ribosomal RNA (X00525)	GAT CGA ATC CGG CCG AGG CGA GGC GCC GCG	48
12	BAR1_52	CCTG 271	EST vb18g06.x1 Soares mouse 3NbMS Mus musculus cDNA clone IMAGE:749338 3' similar to TR:O14862 O14862 INTERFERON-INDUCIBLE PROTEIN (AI661692.1)	GAT CGA ATC CGG CCT GGA CCA CAT CAC GGA	49
20	BAR1_62	GGGC 338	Mus musculus p162 protein mRNA, complete cds (U14172.1)	GAT CGA ATC CGG GGG CCA AGA CGA GGA GGT	50

Table 4 (continued)

Seq ID	Clone ID	Digital Address (MspI)	Database Match (Accession #)	Extended Primer	Primer Seq ID
22	BAR1_67	TAGC 119	EST mm86f09.y1 Stratagene mouse embryonic carcinomaRA (#937318) Mus musculus cDNA clone IMAGE:535337 5' similar to SW:YAQ6_SCHPO Q10106 HYPOTHETICAL 29.7 KD PROTEIN C18G6.06 IN CHROMOSOME I; mRNA sequence (AI325714)	GAT CGA ATC CGG TAG CGG GAA CAC CGA GAA	51
23	BAR1_69	TGCA 213	Mus musculus advillin (Advil-pending), mRNA (NM_009635.1)	GAT CGA ATC CGG TGC AGC ACC GTG AGG TCC	52

We claim:

1. An isolated nucleic acid molecule comprising a polynucleotide chosen from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, and SEQ ID NO:23.

2. An isolated polypeptide encoded by a polynucleotide chosen from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, and SEQ ID NO:23.

3. An isolated nucleic acid molecule comprising a polynucleotide at least 95% identical to the isolated nucleic acid molecule of claim 1.

4. An isolated nucleic acid molecule at least ten bases in length that is hybridizable to the isolated nucleic acid molecule of claim 1 under stringent conditions.

5. An isolated nucleic acid molecule encoding the polypeptide of claim 2.

6. An isolated nucleic acid molecule encoding a fragment of the polypeptide of claim 2.

7. An isolated nucleic acid molecule encoding a polypeptide epitope of the polypeptide of claim 2.

8. The polypeptide of claim 2 wherein the polypeptide has biological activity.

9. An isolated nucleic acid encoding a species homologue of the polypeptide of claim 2.

10. The isolated nucleic acid molecule of claim 1, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the 5'-terminus or the 3'-terminus.

11. A recombinant vector comprising the isolated nucleic acid molecule of claim 1.

12. A recombinant host cell comprising the isolated nucleic acid molecule of claim 1.

13. A method of making the recombinant host cell of claim 12.

14. The recombinant host cell of claim 12 comprising vector sequences.

15. The isolated polypeptide of claim 2, wherein the isolated polypeptide comprises sequential amino acid deletions from either the C-terminus or the N-terminus.

16. An isolated antibody that binds specifically to the isolated polypeptide of claim 2.

17. The isolated antibody of claim 16 wherein the antibody is a monoclonal antibody.

18. The isolated antibody of claim 16 wherein the antibody is a polyclonal antibody.

19. A recombinant host cell that expresses the isolated polypeptide of claim 2.

20. An isolated polypeptide produced by the steps of:

(a) culturing the recombinant host cell of claim 14 under conditions such that said polypeptide is expressed; and

(b) isolating the polypeptide.

21. A method for preventing, treating, modulating, or ameliorating a medical condition, comprising administering to a mammalian subject a therapeutically effective amount of the polypeptide of claim 2 or the polynucleotide of claim 1.

22. The method of claim 21 wherein the medical condition is ataxia telangiectasia.

23. A method for preventing, treating, modulating, or ameliorating a medical condition comprising administering to a mammalian subject a therapeutically effective amount of the antibody of claim 16.

24. The method of claim 23 wherein the medical condition is ataxia telangiectasia.

25. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

(a) determining the presence or absence of a mutation in the polynucleotide of claim 1; and

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or absence of said mutation.

26. The method of claim 25 wherein the pathological condition is ataxia telangiectasia.

27. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising detecting an alteration in expression of a polypeptide encoded by the polynucleotide of claim 1, wherein the presence of an alteration in expression of the polypeptide is indicative of the pathological condition or susceptibility to the pathological condition.

28. The method of claim 27 wherein the alteration in expression is an increase in the amount of expression or a decrease in the amount of expression.

29. The method of claim 27 wherein the pathological condition is ataxia telangiectasia.

30. The method of claim 29 wherein the method further comprises the steps of: obtaining a first biological sample from a patient suspected of having ataxia telangiectasia and obtaining a second sample from a suitable comparable control source;

(a) determining the amount of at least one polypeptide encoded by a polynucleotide of claim 1 in the first and second sample; and

(b) comparing the amount of the polypeptide in the first and second samples;

wherein a patient is diagnosed as having ataxia telangiectasia if the amount of the polypeptide in the first sample is greater than or less than the amount of the polypeptide in the second sample.

31. The use of the polynucleotide of claim 1 or polypeptide of claim 2 for the manufacture of a medicament for the treatment of ataxia telangiectasia.

32. The use of the antibody of claim 16 for the manufacture of a medicament for the treatment of ataxia telangiectasia.

33. A method for identifying a binding partner to the polypeptide of claim 2 comprising:

(a) contacting the polypeptide of claim 2 with a binding partner; and

(b) determining whether the binding partner effects an activity of the polypeptide.

35. The gene corresponding to the cDNA sequence of the isolated nucleic acid of claim 1.

36. A method of identifying an activity of an expressed polypeptide in a biological assay, wherein the method comprises:

(a) expressing the polypeptide of claim 2 in a cell;

(b) isolating the expressed polypeptide;

(c) testing the expressed polypeptide for an activity in a biological assay; and

(d) identifying the activity of the expressed polypeptide based on the test results.

37. A substantially pure isolated DNA molecule suitable for use as a probe for genes regulated in ataxia telangiectasia, chosen from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ

ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, and SEQ ID NO:23.

38. A kit for detecting the presence of the polypeptide of the claim 2 in a mammalian tissue sample comprising a first antibody which immunoreacts with a mammalian protein encoded by a gene corresponding to the polynucleotide of claim 1 or with a polypeptide encoded by the polynucleotide of claim 2 in an amount sufficient for at least one assay and suitable packaging material.

39. A kit of claim 38 further comprising a second antibody that binds to the first antibody.

40. The kit of claim 39 wherein the second antibody is labeled.

41. The kit of claim 40 wherein the label comprises enzymes, radioisotopes, fluorescent compounds, colloidal metals, chemiluminescent compounds, phosphorescent compounds, or bioluminescent compounds.

42. A kit for detecting the presence of a genes encoding an protein comprising a polynucleotide of claim 1, or fragment thereof having at least 10 contiguous bases, in an amount sufficient for at least one assay, and suitable packaging material.

43. A method for detecting the presence of a nucleic acid encoding a protein in a mammalian tissue sample, comprising the steps of:

(a) hybridizing a polynucleotide of claim 1 or fragment thereof having at least 10 contiguous bases, with the nucleic acid of the sample; and

(b) detecting the presence of the hybridization product.

10069025 .021803

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 February 2001 (22.02.2001)

PCT

(10) International Publication Number
WO 01/12647 A1

(51) International Patent Classification[7]: C07H 21/02, 21/04, C12Q 1/68, C12P 19/34, C07K 16/00, A61K 39/395

(21) International Application Number: PCT/US00/22889

(22) International Filing Date: 18 August 2000 (18.08.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/149,868 19 August 1999 (19.08.1999) US

19 Feb02

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/149,868 (CIP)
Filed on 19 August 1999 (19.08.1999)

(71) Applicant (*for all designated States except US*): DIGITAL GENE TECHNOLOGIES, INC. [US/US]; Suite 110, 11149 North Torrey Pines Road, La Jolla, CA 92037 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): BARLOW, Carrolee [US/US]; 6305 El Camino del Teatro, La Jolla, CA 92037 (US). WINROW, Christopher, J. [CA/US]; Apartment 253, 7025 Charmant Drive, San Diego, CA 92122 (US). CALLAHAN, Marie, Lei, A. [US/US]; 13795 Camino Del Suelo, San Diego, CA 92129 (US). PANKRATZ, Daniel, G. [US/US]; 9270-A Regents Road, La Jolla, CA 92037 (US).

(74) Agent: ZIMMERMAN, Roger, P.; McDonnell Boehnen Hulbert & Berghoff, 32nd floor, 300 South Wacker Drive, Chicago, IL 60606 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *With international search report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/12647 A1

(54) Title: GENE EXPRESSION MODULATED IN ATAXIA TELANGIECTASIA TUMORGENESIS

(57) Abstract: Polynucleotides, polypeptides, kits and methods are provided related to regulated genes characteristic of ataxia telangiectasia tumorgenesis.

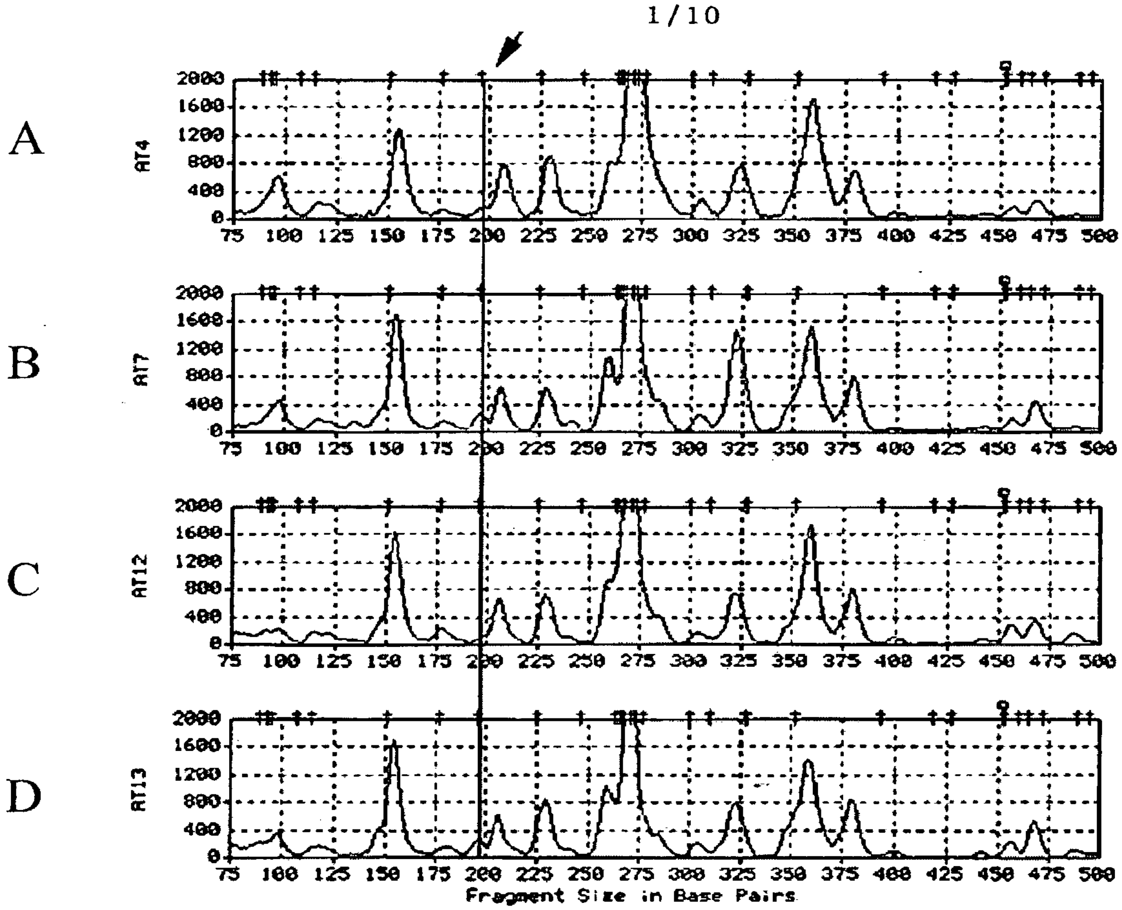
A
B
C
D
AT4
AT7
AT12
AT13
2000
1600
1200
800
400
0
75 100 125 150 175 200 225 250 275 300 325 350 375 400 425 450 475 500
Fragment Size in Base Pairs

Fig. 1

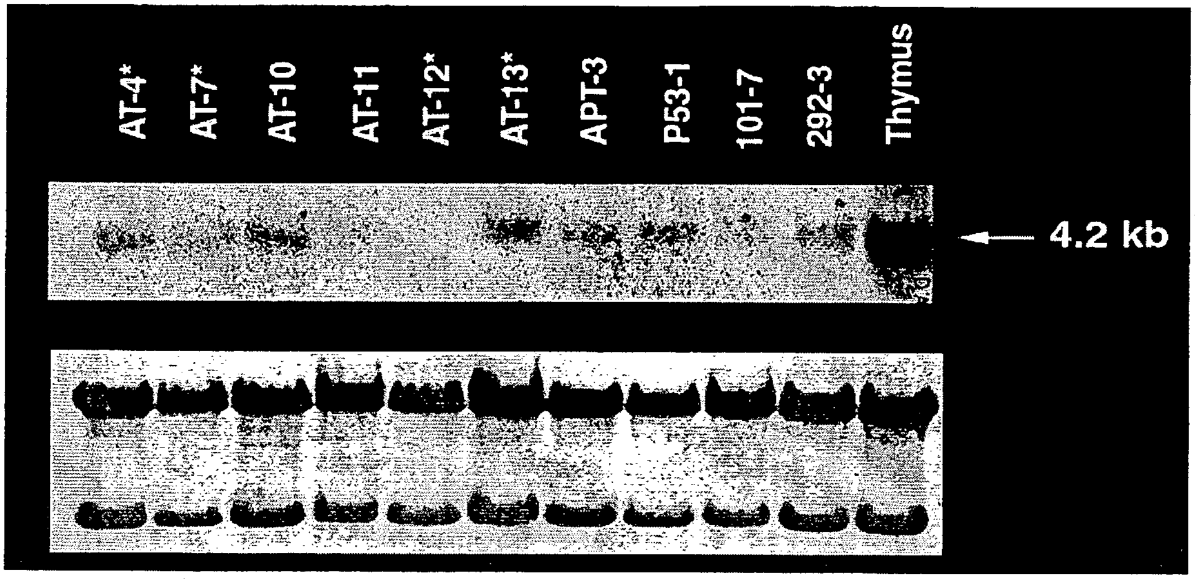

Fig. 2

A
B
C
D

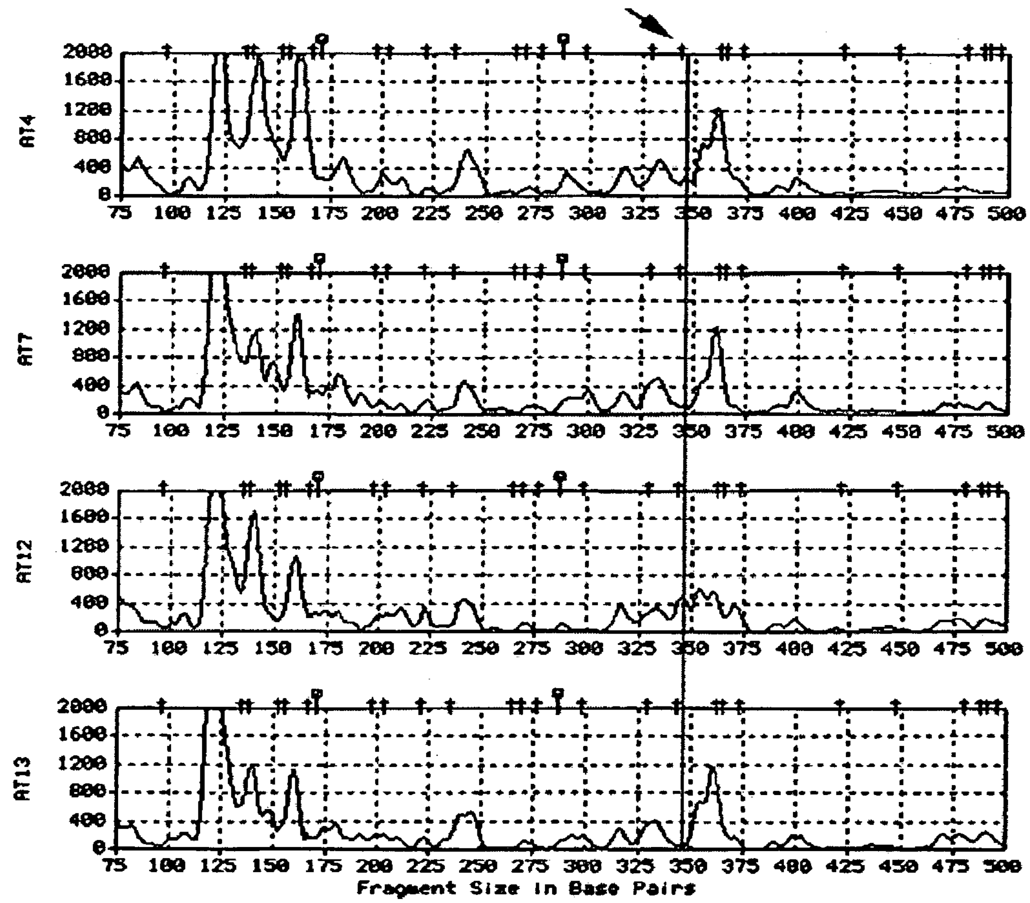
AT4
AT7
AT12
AT13
2000
1600
1200
800
400
0
75 100 125 150 175 200 225 250 275 300 325 350 375 400 425 450 475 500
Fragment Size in Base Pairs

Fig. 3

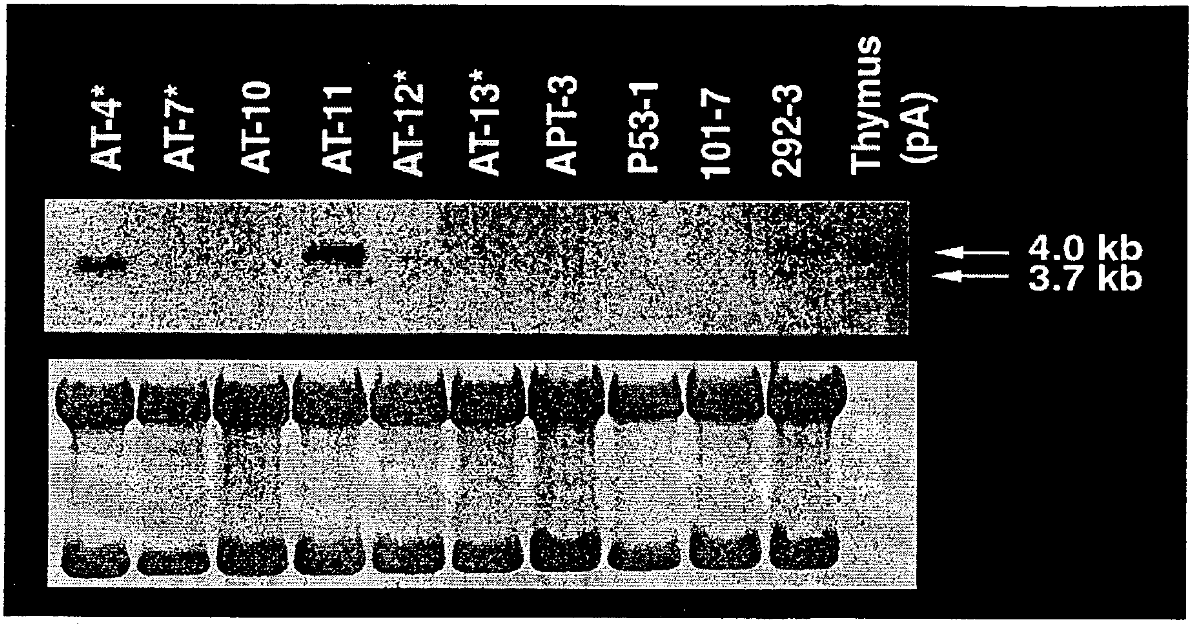
AT-4*
AT-7*
AT-10
AT-11
AT-12*
AT-13*
APT-3
P53-1
101-7
292-3
Thymus (pA)
4.0 kb
3.7 kb

Fig. 4

A
B
C
D

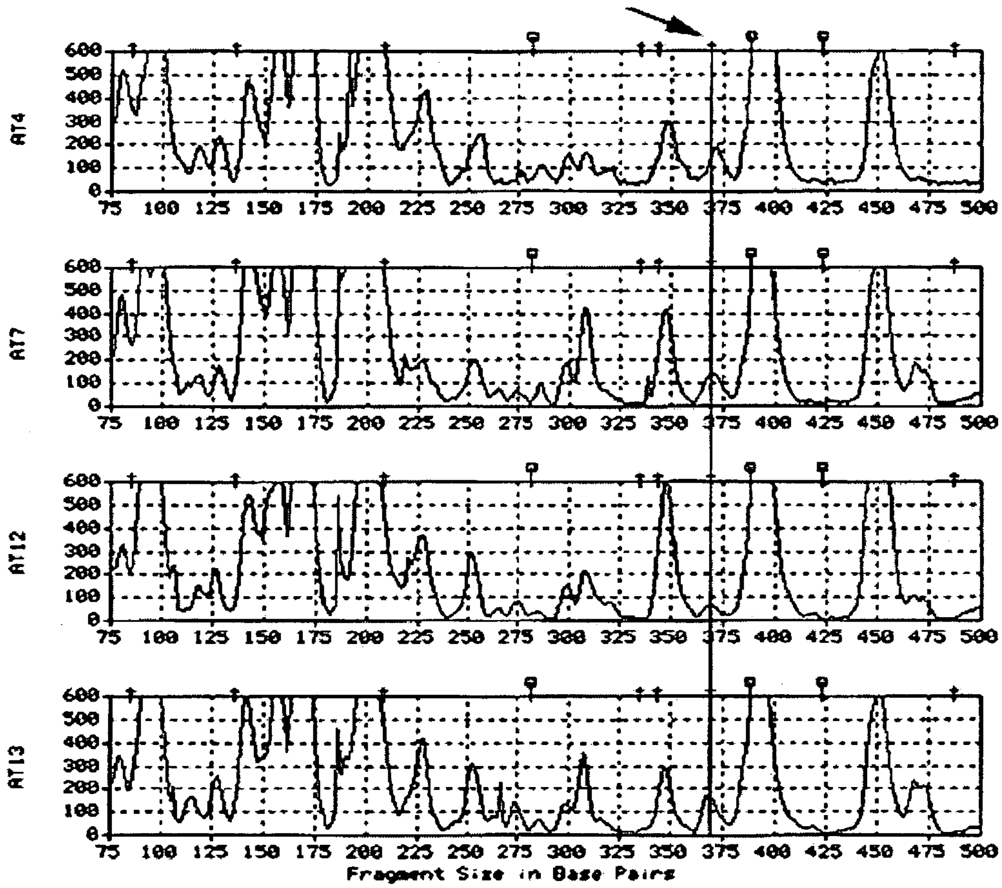
AT4
AT7
AT12
AT13
600
500
400
300
200
100
0
75 100 125 150 175 200 225 250 275 300 325 350 375 400 425 450 475 500
Fragment Size in Base Pairs

Fig. 5

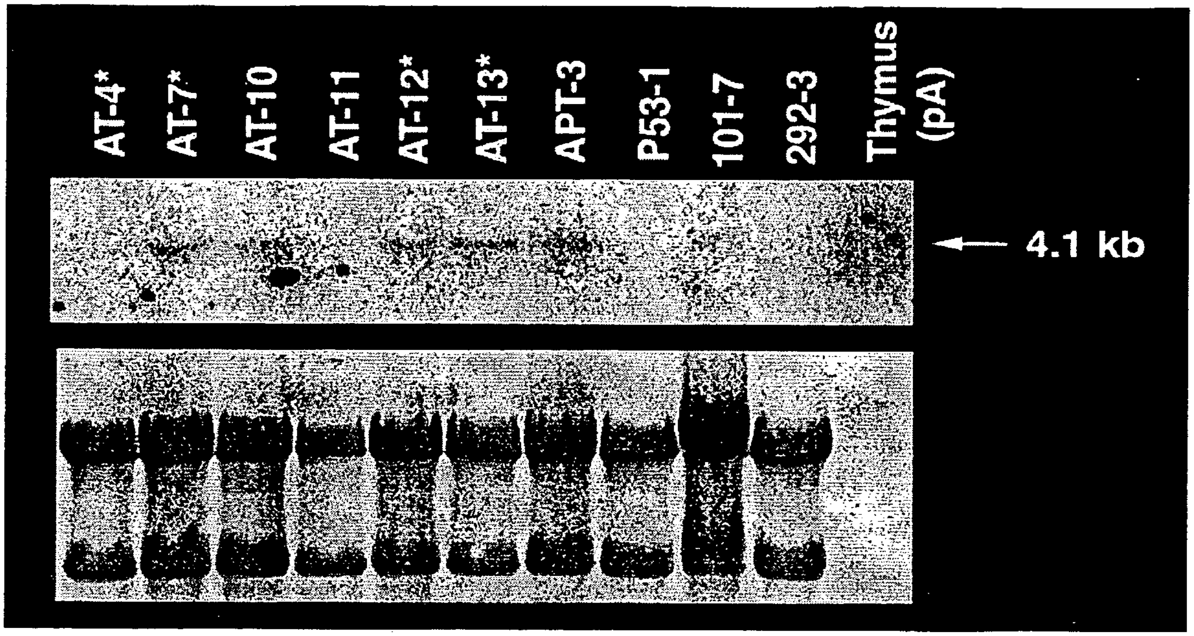
AT-4*
AT-7*
AT-10
AT-11
AT-12*
AT-13*
APT-3
P53-1
101-7
292-3
Thymus (pA)
4.1 kb

Fig. 6

7/10

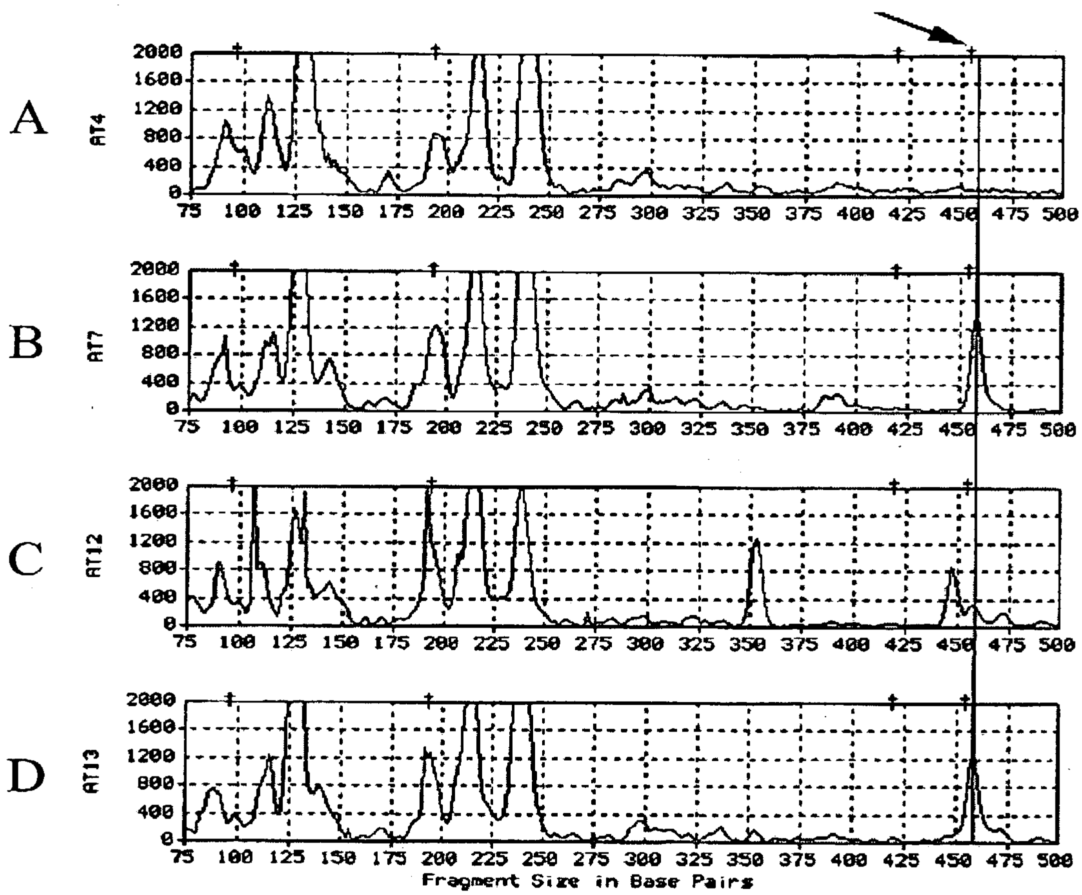
A
B
C
D
AT4
AT7
AT12
AT13
2000
1600
1200
800
400
0
75 100 125 150 175 200 225 250 275 300 325 350 375 400 425 450 475 500
Fragment Size in Base Pairs

Fig. 7

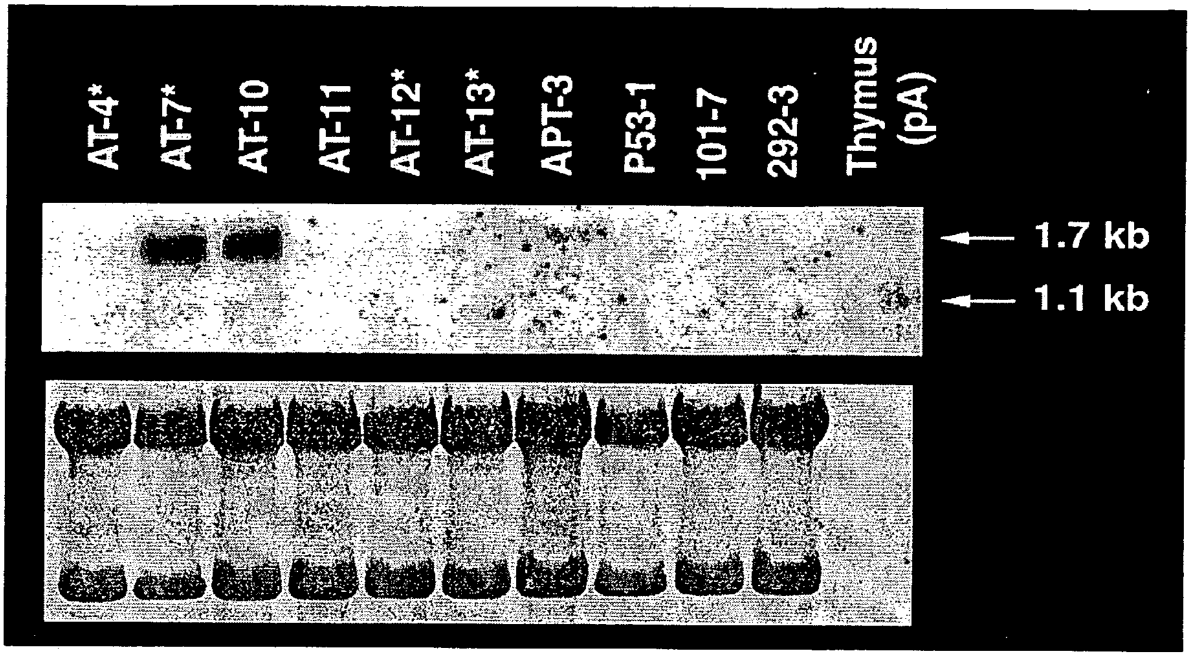
AT-4*
AT-7*
AT-10
AT-11
AT-12*
AT-13*
APT-3
P53-1
101-7
292-3
Thymus (pA)
1.7 kb
1.1 kb

Fig. 8

10/069025

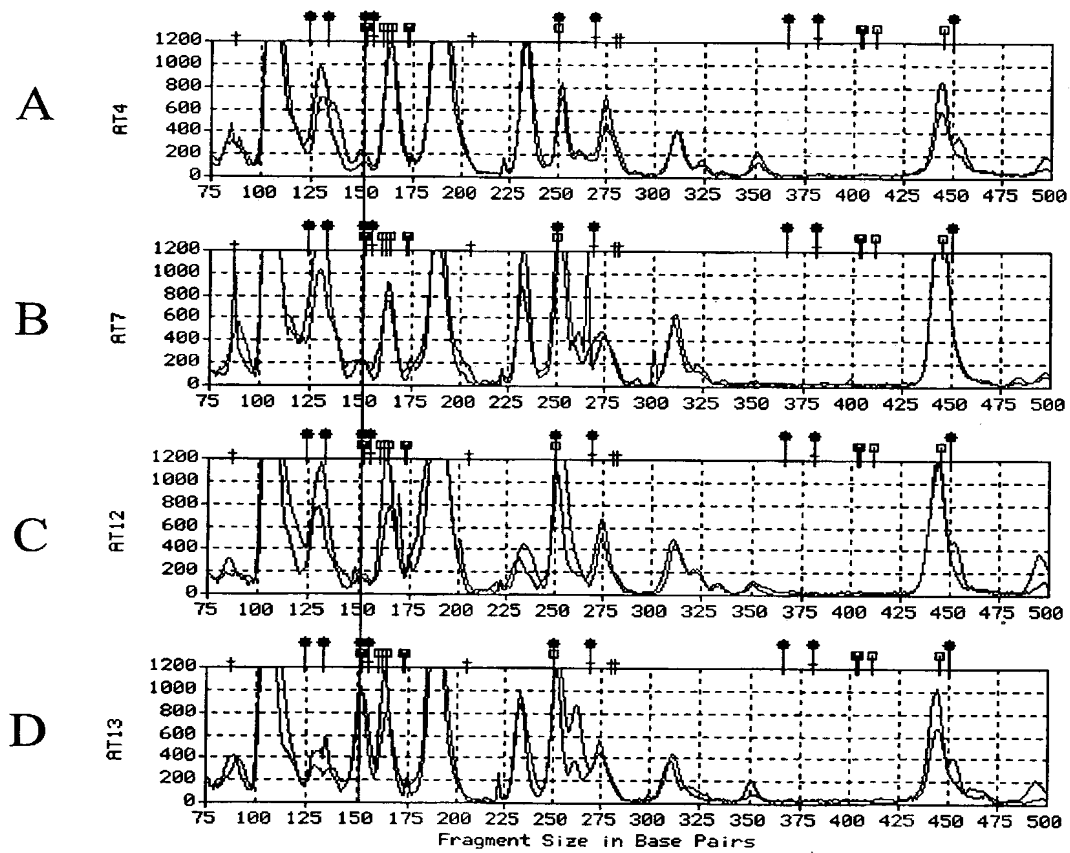

Fig. 9

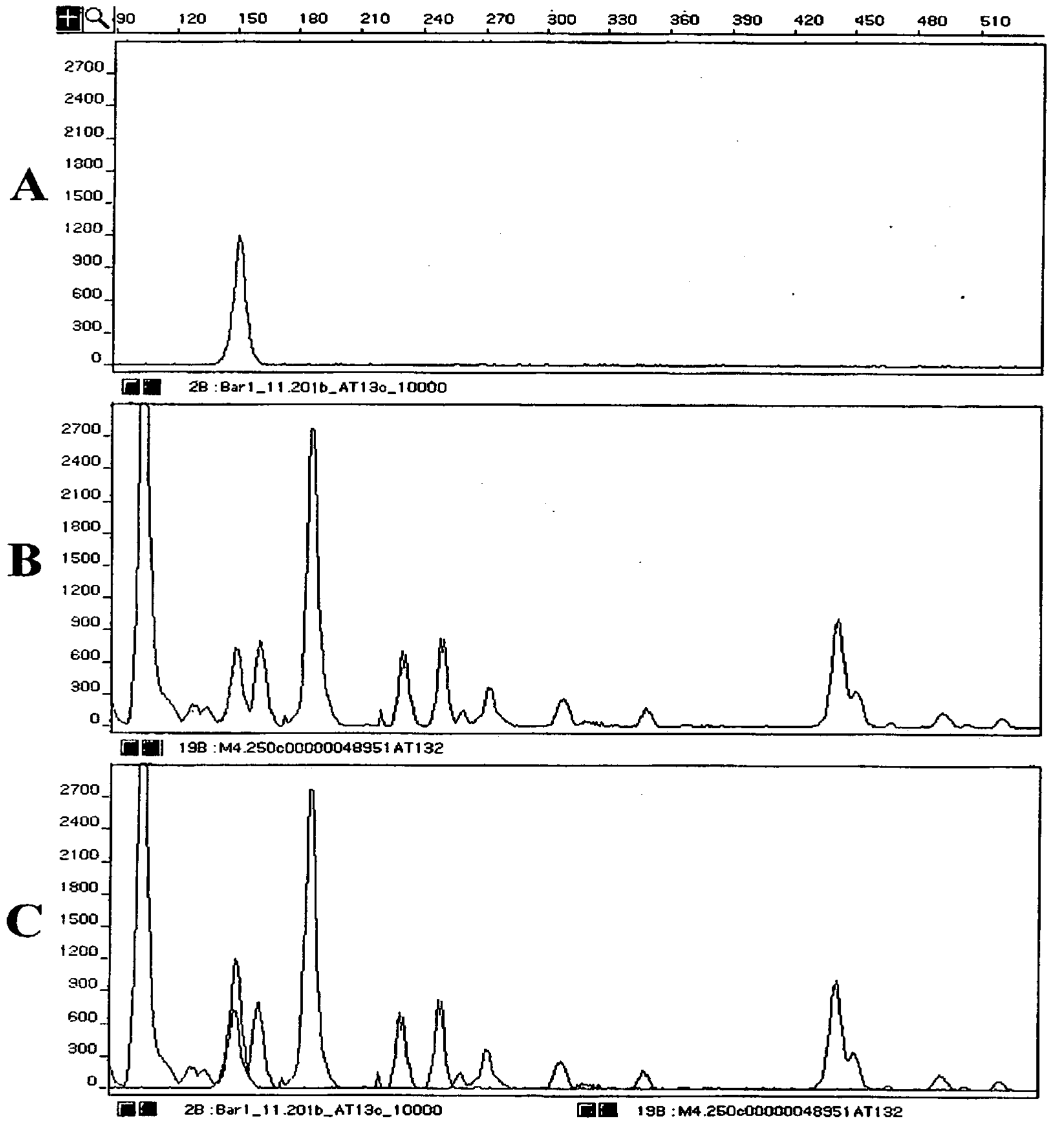
90
120
150
180
210
240
270
300
330
360
390
420
450
480
510
A
B
C
2700
2400
2100
1800
1500
1200
900
600
300
0
2B :Bar1_11.201b_AT13c_10000
19B :M4.250c00000048951AT132

Fig. 10

We hereby claim the benefit under Title 35, United States Code, § 120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, § 112, we acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, § 1.56(a) which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

Application Serial No.	Filing Date
PCT/US00/22889	August 18, 2000

Application Serial No. Filing Date

We hereby declare that all statements made herein of our own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

I hereby appoint:

BERLINER, Brian M., Registration No. 34,549
AJALAT, Christiana M., Registration No. 50,956
FITZSIMMONS, Todd E., Registration No. 44,683
HOLLANDER, Jonathan M., Registration No. 48,717
HSUEH, Peter C., Registration No. 45,574
JAECH, Jonathan A., Registration No. 41,091
9 MILLER, Mark E., Registration No. 31,401
SCHRADER, Scott A., Registration No. 47,928
YAGURA, Ryan K., Registration No. 47,191

all attorneys of the law firm of O'MELVENY & MYERS LLP, 400 South Hope Street, Los Angeles, California 90071-2899, as our attorneys with full powers of substitution and revocation to prosecute this application and to transact all business in the United States Patent and Trademark Office in connection therewith.

Correspondence should be addressed to:

Brian M. Berliner
O'MELVENY & MYERS LLP
400 South Hope Street
Los Angeles, California 90071-2899
Telephone (213) 430-6000

1.00 **Full name of first inventor:** CARROLEE BARLOW

Inventor's signature: Carrolee Barlow

Date: 10-Feb-2003

Residence: 510 Torrey Point Road, Del Mar, California 92014 CA
Citizenship: U.S.A.
Post Office Address: Same as above

2.00 **Full name of second inventor:** CHRISTOPHER J. WINROW

Inventor's signature:

Date: 10-FEB-2003 CA

Residence: 7405 Charmant Drive, Apartment #2105, San Diego, California 92122
Citizenship: ~~U.S.A.~~ CANADA
Post Office Address: Same as above

3.00 **Full name of third inventor:** MARIE LEI A. CALLAHAN

Inventor's signature: Marie Lei A. Callahan

Date: Jan 31, 2003 CA

Residence: 13795 Camino del Suelo, San Diego, California 92129
Citizenship: U.S.A.
Post Office Address: Same as above

4.00 **Full name of fourth inventor:** DANIEL G. PANKRATZ

Inventor's signature:

Date: 2.10.03

Residence: 9270-A Regents Rd., La Jolla, California 92037
Citizenship: U.S.A. CA
Post Office Address: Same as above

216019-67

COMBINED POWER OF ATTORNEY
DECLARATION AND PETITION

As the below named inventors, we hereby declare that:

Our residences, post office addresses and citizenships are as stated below next to our names.

We believe that we are original, first and joint inventor of the subject matter which is claimed and for which a patent is sought on the invention entitled **GENE EXPRESSION MODULATED IN ATAXIA TELANGIECTASIA TUMORGENESIS**, the specification of which was filed with the U.S. Patent and Trademark Office on February 19, 2002, bearing Serial No. 10/069,025.

We hereby state that we have reviewed and understand the contents of the above-identified specification, including the claims.

We acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, § 1.56(a).

We hereby claim foreign priority benefits under Title 35, United States Code, § 119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed: NONE

PRIOR FOREIGN APPLICATION(S)

Number	Country	Date Filed	Priority claimed Yes	Priority claimed No

We hereby claim the benefit under Title 35, United States Code, § 119(e) of any United States provisional application(s) listed below:

Application Serial No.	Filing Date

10/069025

JC10 Rec'd PCT/PTO 19 FEB 2002

SEQUENCE LISTING

<110> Barlow, Carrolee
Winrow, Christopher J
Callahan, Marie Lei A
Pankratz, Daniel G

<120> Gene Expression Modulated in Ataxia Telangiectasia Tumorigenesis

<130> 99-529B

<150> 60/149,868
<151> 1999-08-19

<160> 52

<170> PatentIn version 3.1

<210> 1
<211> 406
<212> DNA
<213> Mus musculus

<400> 1
cggacgggga attcccaaaa acactgcacg aagttaagct gacagtacag aaggatcagg 60

tgtgtgagtc ccagttccaa agttcttaca acagagctaa tgagatatgt gtgggagact 120

caaagatcaa gggagcttcc tttgaggagg attctggagg cccgcttgtg tgtaaaagag 180

cagctgcagg catcgtctcc tacgggcaaa ctgatggatc agctccacaa gtcttcacaa 240

gagttttggg ttttgtatcg tggataaaga aaacgatgaa acacagctaa ctacaagaag 300

caaactagat cctgatctga ccagccatct tccccatagc tgagtccagg attgctctag 360

gacagatggc agcaactgaa taaagaactt tctctgactc caaaaa 406

<210> 2
<211> 54
<212> DNA
<213> Mus musculus

<400> 2
cggcagaggc tcctggcctg cacttttcct attaaagtca ctgacacaca aaaa 54

<210> 3
<211> 101
<212> DNA
<213> Mus musculus

<400> 3
cggccgtgtg tgccttagga gtgggaaact ttgcatttct ctctccttat ccttcttgta 60

agacatccat ttaataaagt ctcatgctga gagagtaaaa a 101

<210> 4
<211> 258

<212> DNA
<213> Mus musculus

<400> 4
cggctcgcag agtgaaggca cccagccatc aggaatgcag ccctgccagc caggcaccat 60

gggtggcagc catcatgctt tcatgaccat tgggcctgct ggtctacctg gtcttagccc 120

aggaagccac caggtaggtt agggtggtca gtgccgagtc tcctgccaga cacagttata 180

cctgcctttc tgcactgctc caggcatgcc cttagagcat ggtgttttaa agctaaataa 240

agtctctaac tacaaaaa 258

<210> 5
<211> 116
<212> DNA
<213> Mus musculus

<400> 5
cgggctcagc actgtctgct cacgtcccca ctgttgcaac acctagtctg tttgcaagga 60

ggacaggaca ggtcacatgg agctttatgt caataaagtc tttatcttgt caaaaa 116

<210> 6
<211> 68
<212> DNA
<213> Mus musculus

<400> 6
cggaacgggg gatgatgggg gattggaaga acttgaactt tttattaaaa tattgtaagt 60

cctaaaaa 68

<210> 7
<211> 172
<212> DNA
<213> Mus musculus

<400> 7
cggaccgcca tgcctggtga gcatatccat cctttgaaag atacactcat gtcctaaacg 60

gagctgtgac tgagatttag tgttttcaat atgaacaaaa agcatttgcc tagcaaatta 120

agttcaaatg aaggcttagt cttttgaaga ataaagcata aaaccatcaa aa 172

<210> 8
<211> 151
<212> DNA
<213> Mus musculus

<400> 8
cggacgctgg agccttcctt aatccacacc tttgaagcca attctcaact ggttgccgtg 60

tacactgatt acatttcagg aggacgtgtt ctatttatgt tgtgcctctg caggcaaagc 120

ccttaataaa tattttgcat actttcaaaa a 151

<210> 9
<211> 300
<212> DNA
<213> Mus musculus

<400> 9
cggacggtgt accccgagga tcgccccagg tggagggaaa gatccaggac caggtcgcgc 60

agcaggagta gaaccccatt tcgcctgtgt gagaaagatc gaatggagct actagaaata 120

gcaaaagcca acgcagcaaa agctctggga acagccaact tcgacttgcc agcaagtctc 180

cgagccaagg aggcaagcca ggggacagct gtttccagca gtgggccaaa ggtggagcat 240

tcagaaaagc agactgaaga tacaactaaa aataccagtg aaaagtcttc tacacaaaaa 300

<210> 10
<211> 43
<212> DNA
<213> Mus musculus

<400> 10
cggagtcccc tcctctgaat taaaatcaca ctgagcggaa aaa 43

<210> 11
<211> 400
<212> DNA
<213> Mus musculus

<400> 11
cggcatcgtc cttgttaagc actgctttcc agaccgtgtc tgacttagtc aacctagagg 60

attttatagt ttcactctaa tgaagtcctg attaacacag agcagcggag cagcggtccc 120

tcccctgagt ggcagcagca ggaagccccg tttgtaatgt gcgttgcgct gctttaccgt 180

acacccagtg gcattctttt tactataacg gttttccccc ctctcagaag agaaatctta 240

aaatttgcaa cagacctcct aacttttgtt tacttttttt gtcttaacga tagatttgaa 300

aatgaaaaat gttaaaaaga caactggatc tgttgtccgt attatatttg caatctgaaa 360

tttgtgtgaa ttgatttagt aaaatgttaa actggaaaaa 400

<210> 12
<211> 219
<212> DNA
<213> Mus musculus

<400> 12
cggcctggac cacatcacgg aggaagaact gaaacggttc aagtactttg ctttgactga 60

gtttcagatt gccaggagca cactcgacgt cgcagatagg acagaattag ctgaccacct 120

gattcaaagt gcatgtgcgg cgtctggagt gaccaaggcc attaatattt tccagaagtt 180

gaattatatg catattgcaa atgctcttga agagaaaaa 219

<210> 13
<211> 89
<212> DNA
<213> Mus musculus

<400> 13
cgggctcttt gcacagttct aacctataca acttggtctg ctcctcctac caacttattt 60

aatagacatt aaagtaaaac gcccaaaaa 89

<210> 14
<211> 42
<212> DNA
<213> Mus musculus

<400> 14
cggctatcag tgctaacaga aaacattgta gaaggccaaa aa 42

<210> 15
<211> 247
<212> DNA
<213> Mus musculus

<400> 15
cggaccatcc ctgctctgtg tccatctact catccaatgg agtcattagc agtcaatcgc 60

agcttcactg gacactgagg ggtcacagac ttaggctccc tttgggtcac gtccagcgtg 120

tcctagactt tatcatcttt cagaggcggt aggcagactg ttcacaaagg ctttctctag 180

ctttccactg caattaatct tggtcactcc ctcaaataag tatatttttt ctagaaaagg 240

ggaaaaa 247

<210> 16
<211> 260
<212> DNA
<213> Mus musculus

<400> 16
cggaccctga ttcgcatcat ggtgtctcgc ggcgagcttg acctgctgga tatccgagca 60

gagtataagc ggatgtacgg caagtcactg taccacgata tcacgggaga tacttctgga 120

gactatcgaa agattccact gaagatctgt ggtggcaacg actgagctgt ggcagtggct 180

cacctgtgcc ccctgctggc attggccagg aaacgtccca aagagcacac ttcggtttcc 240

ctatctaata accccaaaaa 260

<210> 17
<211> 209
<212> DNA
<213> Mus musculus

<400> 17
cggaccgtgc aaaggtagca taatcacttg ttccttaatt agggactagc atgaacggct 60

aaacgagggt ccaactgtct cttatcttta atcagtgaaa ttgacctttc agtgaagagg 120

ctgaaatata ataataagac gagaagaccc tatggggctt aaattatata acttatctat 180

ttaatttatt aaacctaatg gcccaaaaa 209

<210> 18
<211> 385
<212> DNA
<213> Mus musculus

<400> 18
cggagtggac agcacaagcc atggaagcgc agatggccca gtgagccctc agcgctgtgc 60

gcactgcatc tttgccactt cgtgcagatt gattcactgc ccctcaaagc atttgcatca 120

tgaccttaga catctcagcc tcctttgctg gatcccactt gaagaagatg ctagagcaag 180

aagtgcatcg tttaaacata atagctcccg tgagatctgg gtctgtgtat tgaactaact 240

cagacgcttt gaaaacttgt tctttaagtg gaagtgaaag acctgaagtt tgggatttgt 300

gatcactttg tagtcatgta aaacttaagt gctttgtgcc tctccacatg tggttactct 360

aataaacgga gcggtgagcc aaaaa 385

<210> 19
<211> 83
<212> DNA
<213> Mus musculus

<400> 19
cggccgaggc gaggcgccgc gcggaaaacc gcggcgggtg ttgacgcgat gtgatttctg 60

cccagtgctc tgaatgtcaa aaa 83

<210> 20
<211> 286
<212> DNA
<213> Mus musculus

<400> 20
cgggggccaa gacgaggagg tgctgatgat gaacggtcat cctggcggaa tgctgatgat 60

gataggggtc ccaggagagg catggatgat gataggggtc ccaggcgagg gctggatgat 120

gaccgaggac cttggaggaa tgctgctgaa gataggattt ccaggagagg tgcagatgat 180

gaccgagggc cttggagaaa tatggatgat gatagggttc ctagaagggg tgatgatgca 240

agacctggtc cttggagatc atttgtcgag ccaggtggat gaaaaa 286

<210> 21

<211> 124
<212> DNA
<213> Mus musculus

<400> 21
cggtaccaat gtatagctgt gttggagaaa tatggaagat aaactataca ggatatttgc 60

acattctcat ctgtctctct ggtgaccgat aattcagttt gctccagctg cccaagccca 120

aaaa 124

<210> 22
<211> 68
<212> DNA
<213> Mus musculus

<400> 22
cggtagcggg aacaccgaga acgaagccag cctggctttc gaaaacgcct gggtttgctg 60

gagaaaaa 68

<210> 23
<211> 160
<212> DNA
<213> Mus musculus

<400> 23
cggtgcagca ccgtgaggtc cagtaccacg agtccgacac cttccgtggc tacttcaagc 60

agggcatcat ctataaaaag gggggtgtgg cctctgggat gaagcacgtg gagaccaaca 120

cctatgatgt gaagaggctg ctacacgtaa aggggaaaaa 160

<210> 24
<211> 48
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic primer

<220>
<221> misc_feature
<222> (47)..(48)
<223> N stands for any nucleotide.

<220>
<221> misc_feature
<222> (46)..(46)
<223> V stands for A, G, or C.

<400> 24
gaattcaact ggaagcggcc gcaggaattt tttttttttt tttttvnn 48

<210> 25

<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic primer

<400> 25
aggtcgacgg tatcgg   16

<210> 26
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic primer

<220>
<221> misc_feature
<222> (16)..(16)
<223> N stands for any nucleotide.

<400> 26
ggtcgacggt atcggn   16

<210> 27
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic primer

<400> 27
gagctccacc gcggt   15

<210> 28
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic primer

<220>
<221> misc_feature
<222> (13)..(16)
<223> N stands for any nucleotide.

<400> 28
cgacggtatc ggnnnn   16

<210> 29
<211> 16

<212>   DNA
<213>   Artificial Sequence

<220>
<223>   Description of Artificial Sequence: synthetic primer

<400>   29
cgacggtatc gggagc   16

<210>   30
<211>   30
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   Description of Artificial Sequence: Extended TOGA primer for BAR1
        _7.

<400>   30
gatcgaatcc ggacggtgta ccccgaggat   30

<210>   31
<211>   30
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   Description of Artificial Sequence: Extended TOGA primer for BAR1
        _8.

<400>   31
gatcgaatcc gggctctttg cacagttcta   30

<210>   32
<211>   30
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   Description of Artificial Sequence: Extended TOGA primer for BAR1
        _11.

<400>   32
gatcgaatcc ggccgtgtgt gccttaggag   30

<210>   33
<211>   30
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   Description of Artificial Sequence: Extended TOGA primer for BAR1
        _16.

<400>   33
gatcgaatcc ggaacggggg atgatggggg   30

<210> 34
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1 _20.

<400> 34
gatcgaatcc ggtaccaatg tatagctgtg 30

<210> 35
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1 _21.

<400> 35
gatcgaatcc ggcagaggct cctggcctgc 30

<210> 36
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1 _22.

<400> 36
gatcgaatcc ggctatcagt gctaacagaa 30

<210> 37
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1 _24.

<400> 37
gatcgaatcc gggctcagca ctgtctgctc 30

<210> 38
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1 _27.

<400> 38
gatcgaatcc ggacggggaa ttcccaaaaa   30

<210> 39
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1
       _28.

<400> 39
gatcgaatcc ggctcgcaga gtgaaggcac   30

<210> 40
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1
       _29.

<400> 40
gatcgaatcc ggagtcccct cctctgaatt   30

<210> 41
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1
       _34.

<400> 41
gatcgaatcc ggaccatccc tgctctgtgt   30

<210> 42
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1
       _35.

<400> 42
gatcgaatcc ggaccctgat tcgcatcatg   30

<210> 43
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1 _37.

<400> 43
gatcgaatcc ggaccgccat gcctggtgag   30

<210> 44
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1 _38.

<400> 44
gatcgaatcc ggaccgtgca aaggtagcat   30

<210> 45
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1 _39.

<400> 45
gatcgaatcc ggacgctgga gccttcctta   30

<210> 46
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1 _41.

<400> 46
gatcgaatcc ggagtggaca gcacaagcca   30

<210> 47
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1 _49.

<400> 47
gatcgaatcc ggcatcgtcc ttgttaagca   30

<210> 48
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1 _50.

<400> 48
gatcgaatcc ggccgaggcg aggcgccgcg 30

<210> 49
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1 _52.

<400> 49
gatcgaatcc ggcctggacc acatcacgga 30

<210> 50
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1 _62.

<400> 50
gatcgaatcc gggggccaag acgaggaggt 30

<210> 51
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1 _67.

<400> 51
gatcgaatcc ggtagcggga acaccgagaa 30

<210> 52
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Extended TOGA primer for BAR1 _69.

<400> 52
gatcgaatcc ggtgcagcac cgtgaggtcc 30